



Искусственный Интеллект

Artificial intelligence - AI

Материалы внутренних семинаров компании Nihol

Искусственный Интеллект – что это?

Искусственный интеллект (ИИ) по состоянию на 2023 год - один из самых популярных, обсуждаемых и одновременно противоречивых и многозначных терминов. Рынок искусственного интеллекта оценивается консалтинговыми компаниями в сотни миллиардов долларов.



Интеллект (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. Интеллект — это общая способность к познанию и решению трудностей, которая объединяет все познавательные способности человека: ощущение, восприятие, память, представление, мышление, воображение.

Многие авторы отмечали, что словосочетание «Искусственный интеллект» – это неудачный термин, который тем не менее стал общеупотребимым. Целый ряд специалистов предлагали свои уточнения этого названия. Например, профессор Константин Воронцов, отметил, что ИИ логичнее было бы расшифровать как «имитация интеллекта».



Искусственный интеллект (ИИ, англ. Artificial intelligence, AI) — наука и технология создания интеллектуальных машин, особенно интеллектуальных компьютерных программ. ИИ связан со сходной задачей использования компьютеров для понимания человеческого интеллекта, но не обязательно ограничивается биологически правдоподобными методами.

Дать однозначную исчерпывающую и согласованную дефиницию ИИ очень сложно, поэтому и существуют сотни частных определений. Почему же термин, которым люди пользуются почти 70 лет, до сих пор понимается разными группами специалистов по-разному? Какие определения ИИ существуют и чем они различаются?

Искусственный Интеллект – определения разных лет

Десять определений ИИ, предложенных разными авторами за последние 50 лет:

	Определение	Автор, год
1	Автоматизация задач, считающиеся человеческими: мышление, принятие решений, решение проблем, обучение и т. д.	Bellman, 1978
2	Область исследований - разъяснение и эмуляция разумного поведения в терминах вычислительных процессов	Schalkoff, 1990
3	Искусство создания машин, выполняющих функции, которые требуют интеллекта, как если бы их выполняли люди	Курцвейл, 1990
4	Наука и инженерия создания интеллектуальных машин	McCarthy, 2007
5	Область, изучающая компьютерные технологии, которые делают возможным восприятие, рассуждения и действия машин на их основе	Winston, P. H., 1992
6	Деятельность, направленная на создание интеллектуальных машин, а интеллект – это качество, которое позволяет объекту функционировать в окружающей среде надлежащим образом и с предвидением	Nilsson, N. J., 2010
7	Зонтичный термин, охватывающий целый ряд алгоритмов, позволяющих оптимизировать поиск в Интернете, нацеливать рекламу, утверждать потребительские кредиты и направлять водителей	Эндрю Нг, 2017
8	Выполнение компьютером действий, для которых обычно требуется человеческий интеллект	Эми Вебб, 2017
9	Интеллектуальная деятельность, которая раньше выполнялась только на основе интеллекта человека, а теперь может быть выполнена компьютером, включая распознавание речи, машинное обучение и обработку естественного языка	Вишал Сикка, 2017
10	Комплекс технологических решений, позволяющий имитировать <u>когнитивные</u> 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.»	Нац. Стратегия развития ИИ в РФ до 2030 г.

Искусственный Интеллект – миф или реальность?

Некоторые эксперты утверждают, что уровень развития технологий искусственного интеллекта в стране – это важнейший критерий технологической, экономической и военно-стратегической мощи государства, показатель её конкурентоспособности.

Другие специалисты, утверждают, что ИИ – это миф, никакого искусственного интеллекта пока не создано! Например:



«На самом деле искусственного интеллекта для бизнеса еще не существует. Перефразируя Марка Твена (или, скорее, распространенное неверное цитирование того, что на самом деле сказал Твен), сообщения о рождении ИИ были сильно преувеличены.»

Многие компании заявляют, что используют его, но они обманывают себя — и вас тоже. ИИ лучше всего рассматривать как набор технологий следующего поколения, которые бизнес еще не начал использовать.

Итак, если это не ИИ, то какие технологии сегодня используют предприятия?

Для многих это автоматизация. Организации используют процессы, которые существовали десятилетиями, но выполнялись людьми от руки (например, ввод информации в книги) или в электронных таблицах. Теперь эти же процессы переводятся в код для машин. Машины, как пианино, бездумно выполняют действия, которых не понимают.

Когда появится искусственный интеллект для бизнеса, ожидайте радикальной трансформации. Это породит организации, построенные иначе с нуля. Вместо того, чтобы спрашивать, какие продукты или услуги могут предоставлять люди и как машины могут помочь им в этом, руководители этих предприятий начнут с того, что может сделать искусственный интеллект, и построят бизнес вокруг этого. Настоящий ИИ может даже создать компанию самостоятельно. Это будет новая парадигма.»

Источник: Scientific American. «AI Doesn't Actually Exist Yet». By Max Simkoff, Andy Mahdavi on November 12, 2019

Искусственный Интеллект – почему сложно дать определение?

Во-первых, ИИ быстро развивается, появляется все больше новых решений, меняется наполнение термина, и трудно дать толкование, которое было бы достаточно гибким, охватывая все новые подходы к реализации ИИ.

Во-вторых, ИИ опирается на исследования разных наук, как естественных, так и гуманитарных (компьютерные науки, статистика, психология, нейробиология, философия), каждая из которых имеет свой понятийный аппарат, свои взгляды на предмет, и не всегда эти взгляды совпадают. Часть философов считают, что сознание может существовать только в рамках живой природы, другие полагают, что интеллект и наличие волеполагания, жажда познания и даже любовь – это все атрибуты, которые ИИ может приобрести на определенной стадии своего развития.

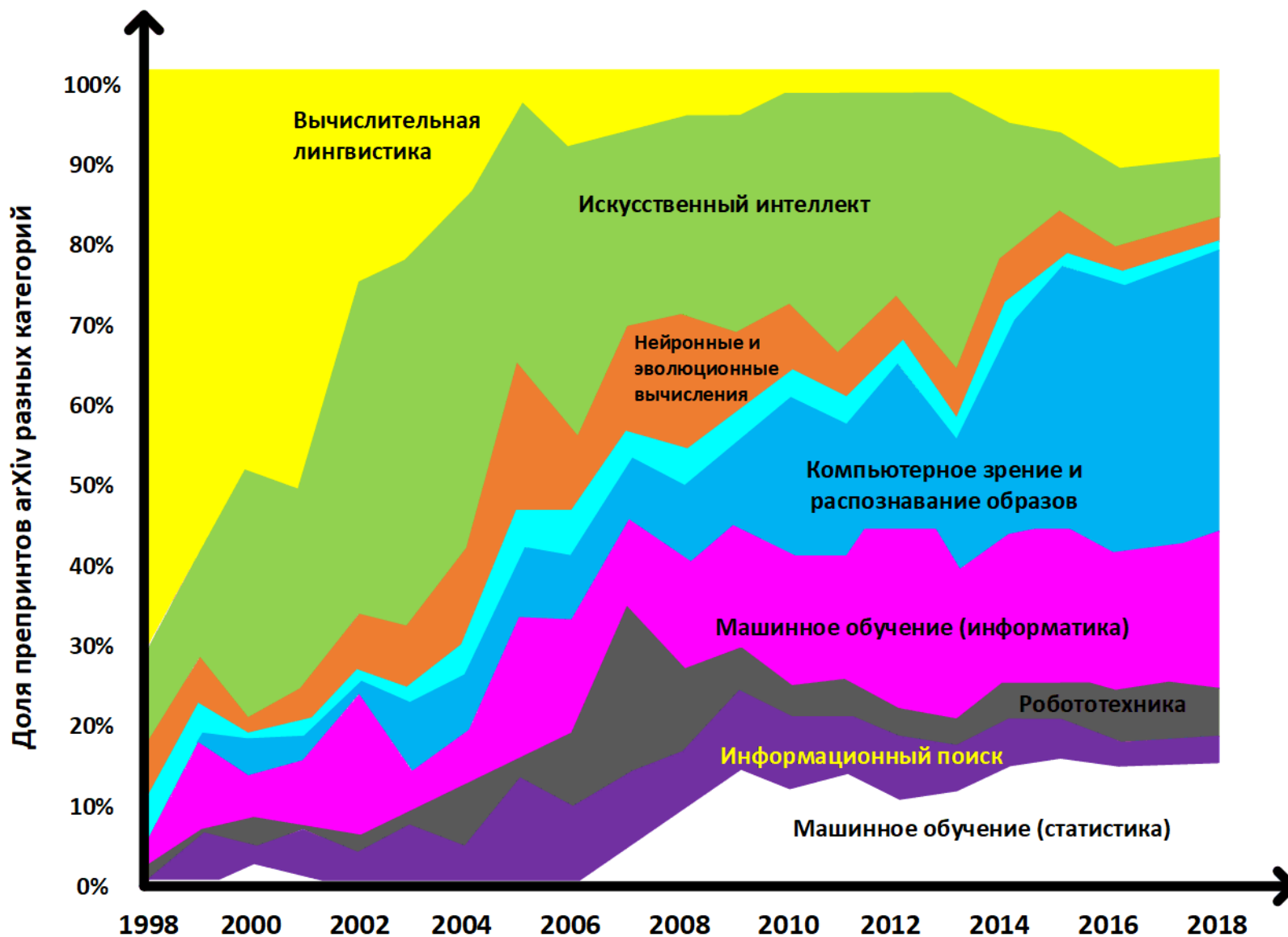
Представители разных школ и учений (материалисты и идеалисты, сторонники эволюционной теории и ее противники и т. п.) не могут достигнуть консенсуса в толковании термина ИИ, особенно в вопросах, затрагивающих философские аспекты бытия.

С философской точки зрения термин интеллект привязан к биологическому объекту, а вопрос о том, могут ли неживые объекты обладать ощущением не решен. Поэтому вопрос о том, может ли машина обладать разумом и, соответственно, интеллектом, пытаются отложить на будущее.

Разные участники сообщества ставят перед собой разные задачи. Для одних исследование ИИ связано с узко научными целями, например, как с помощью компьютерного моделирования понять механизмы работы интеллекта человека. Для других целью является создание умных машин, которые позволят решать новые научные и практические задачи, не подвластные сегодня человеку. Для третьих задача формулируется как бизнес и состоит в использовании новых ИИ-технологий везде, где их использование несет экономическую выгоду.

В-третьих, термином ИИ могут назвать разработки самого разного уровня сложности, опирающиеся на разные направления искусственного интеллекта, разные технологии, реализованные как в виртуальной среде, так и воплощенные в физическое устройство (в виде умной робототехники).

Искусственный Интеллект – почему сложно дать определение?



Большая часть авторов, базируются на утверждении, что ИИ – это некоторая область деятельности человека. Другая часть экспертов делает акцент на том, что ИИ – это определенное свойство машин.

Это принципиально разные подходы. ИИ, как наука, может заниматься чисто теоретическими вопросами. Ответ на вопрос «Что такое ИИ как наука?» предполагает в первую очередь обозначение перечня областей исследования, научных школ, направлений и т. п.

Анализируя долю статей в области различных дисциплин, можно судить об изменении фокуса научных изысканий в области искусственного интеллекта в разные периоды времени.

Искусственный Интеллект – виды и подкатегории

Во избежание путаницы было введены дополнительные термины для распределения ИИ на подкатегории в соответствии с его возможностями. Всего выделили 3 вида ИИ:

1. Слабый ИИ (или же узкий, прикладной) - ANI, Artificial Narrow Intelligence;
2. Сильный ИИ (или общий) - AGI, Artificial General Intelligence
3. Супер ИИ (Super AI).

Узкий ИИ применяет свой “интеллект” к ограниченному кругу задач. Сюда можно отнести виртуальных ассистентов Siri и Google Assistant, рекомендательные системы в YouTube и Netflix, переводчик Google Translator. По сути, все технологии, которые сегодня носят название искусственного интеллекта, было бы целесообразно называть слабым ИИ, так как сфера их применения довольно узкая. Они не могут мыслить креативно, придумывать что-то новое, действовать спонтанно.

Слабому ИИ противопоставляют Общий ИИ. Одно из определений Общего ИИ звучит так: это “система, способная применять интеллект к множеству непредусмотренных типов проблем”. Среди людей такое качество наверное можно назвать гибкостью ума, многозадачностью, или креативностью. То есть, общий ИИ будет соответствовать человеческому уровню интеллекта. Он не будет нуждаться в инструкциях и поддержке со стороны человека.

Супер ИИ это система с полным самосознанием и сформированным мышлением, превосходящим человеческое. - “интеллект, который намного умнее, чем лучший человеческий мозг, практически во всех областях, в том числе в научном творчестве, общей мудрости и социальных навыках” (Ник Бостром, University of Oxford). Предположительно Super AI сможет самостоятельно перепрограммироваться, создавать системы нового направления и алгоритмы без вмешательства человека.

Существуют предсказания, что общий ИИ со способностью к обучению будет достигнут в 2030 году. В какой-то момент он неизбежно превзойдет человека по уровню интеллекта, и уже к 2045 году будет достигнут уровень Супер ИИ. Но для того, чтобы это произошло, сначала должен появиться компьютер с умением учиться — этим занимается подраздел искусственного интеллекта под названием “машинное обучение”.

Существующий ИИ – слабый ИИ?

Сравнительно недавно, говоря об отличии общего и узкого ИИ авторы указывали на то, что узкий (слабый) ИИ – это интеллект, способный выполнять конкретную интеллектуальную функцию или их ограниченный набор подчас лучше, чем человек. При этом, превышая возможности человека в узкой области, такая система, как правило, не имеет интеллектуальных способностей в других сферах, в отличие от человека, который обучается решению задач в самых разных областях.

Однако в 2023 году – с появлением мультимодальных моделей – это отличие звучит не так четко. ИИ по-прежнему не обладает возможностью обучения по столь широкому спектру направлений, как человек, но, например, мультимодальная нейросеть Gato (от Исп. Кошка) способна выполнять 604 типа задач, включая создание описаний к изображениям, ведение диалога, укладку блоков с помощью роборуки, игру в аркадные игры, выполняя при этом более половины из перечисленных задач лучше среднего человека. А чат бот ChatGPT от корпорации OpenAI умеет отвечать на вопросы, искать информацию почти как Google, писать забавные диалоги и рассказы, составлять электронные письма и метатеги, разрабатывать простые приложения (писать фрагменты программного кода).

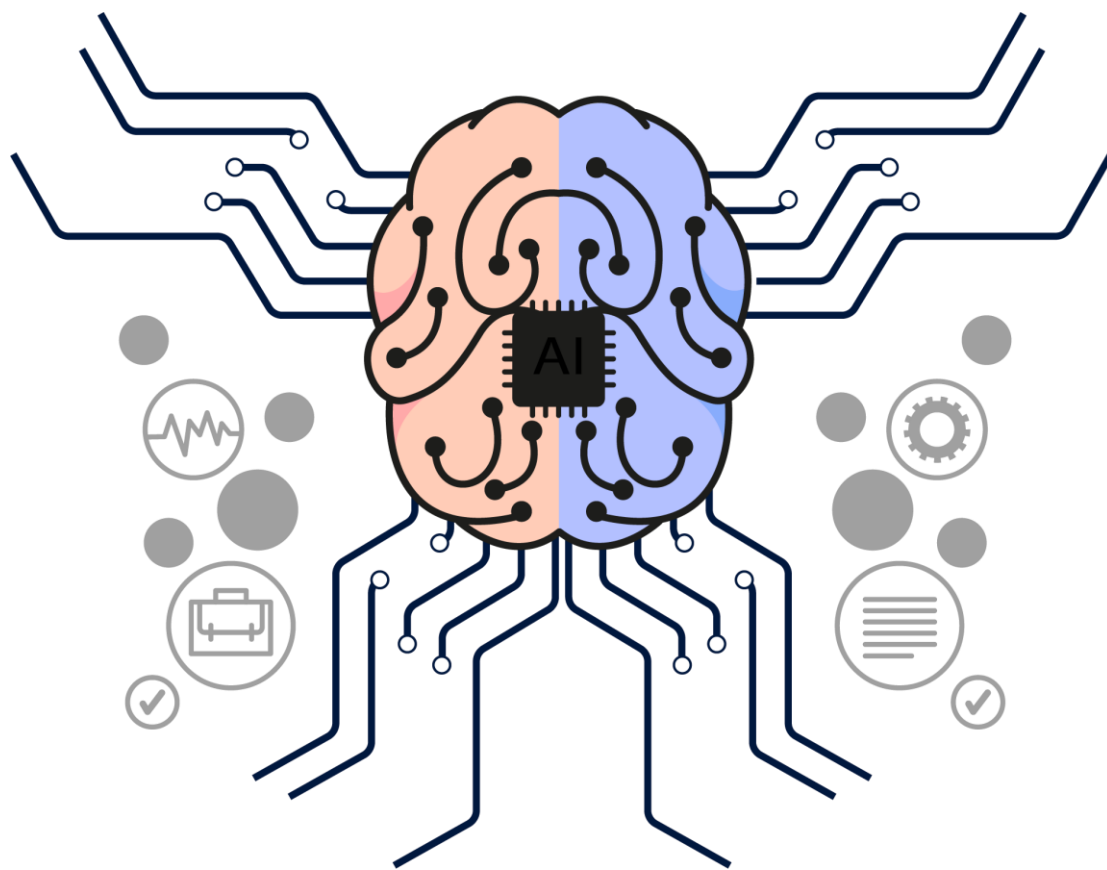
Появление каждой новой перспективной возможности дает повод энтузиастам говорить о том, что пройден еще один рубеж на пути создания сильного ИИ. При этом скептики продолжают говорить о том, что ИИ никогда не будет обладать широтой когнитивных способностей человека, не будет обладать эмоциями, самосознанием и целеполаганием.

Эти рассуждения часто обосновывают тем, что у машины, соизмеримой по интеллекту с человеком, должны быть цели и мотивы, подобные тем, которые есть у человека. Цели и мотивы поведения человека зависят от многих факторов. Некоторые из них обусловлены нашими инстинктами, например, желание секса, еды и крова. На поведение человека и его цели оказывают влияние эмоции – такие, как страх, гнев или ревность.

Некоторые наши цели и мотивы носят общественный характер, например люди подвержены воздействию таких коллективных понятий как «успешный человек».

Существующий ИИ – слабый ИИ?

«Революционный поворот в понимании того, что индивидуальным может быть только мозг, а разум – исключительно коллективный феномен, кардинально изменит траекторию исследований общего (сильного) искусственного интеллекта. Его исследователям предстоит найти новые, принципиально иные архитектуры и алгоритмы, ориентированные уже не только на нейронные, а на социокогнитивные гиперсети»

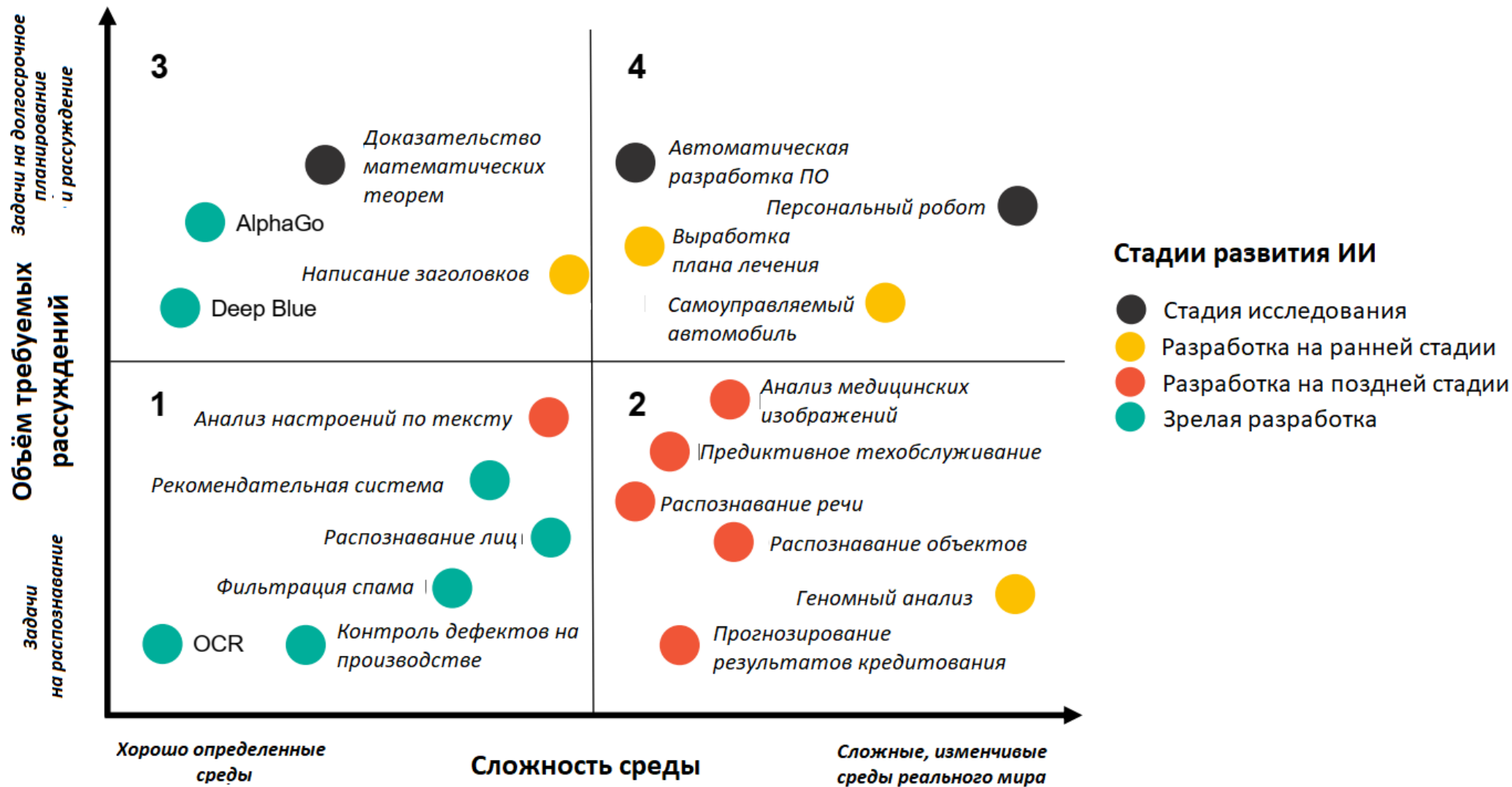


Источник: «Cognitive Neuroscience Meets the Community of Knowledge» Арон Барби, Ричард Паттерсон, Стивен Сломан).

В рамках описанной выше терминологии все существующие на сегодняшний день системы ИИ относятся к категории слабого ИИ.

Следует отметить, что термины «слабый» и «узкий» отражают разные аспекты понятия. Слово «слабый» не совсем уместно, когда мы, например, говорим о программах, обыгрывающих человека в ту или иную игру, здесь скорее более подходит термин «узкий», там, где ИИ не достигает уровня человека, более уместен термин «слабый».

Существующий ИИ – примеры ИИ приложений



Возможен ли Сильный ИИ?

В отличие от слабого, сильный ИИ, или интеллект общего назначения, – это гипотетическая машина будущего, которая будет способна понимать мир на уровне, сопоставимом с уровнем понимания его человеком, и обучаться выполнению всего спектра интеллектуальных задач, которые может выполнить человек.

Подобное деление помогает терминологически разделить два разных понятия и раскрыть одну из основных причин отсутствия консенсуса в определении ИИ. Действительно, часть специалистов (в большей степени гуманитарии) склонны воспринимать термин ИИ как сильный ИИ, как интеллектуальную машину, наделенную свойствами сознания, осознания, мышления, чувственного восприятия действительности, волеизъявления.

Другая часть сообщества, занятая разработкой технологий ИИ и различного рода приложений на его основе, привыкла называть искусственным интеллектом именно то, с чем они имеют дело (то есть со слабым ИИ). И в этой связи понимание под термином ИИ-решений, относящихся к категории «слабый ИИ», не вызывают у них отторжения. Несмотря на то, что «сильный ИИ» и «слабый ИИ» существенно утоняют понятие, люди в основном пользуются термином ИИ, предполагая, что в зависимости от контекста понятно, о чем идет речь.

Подчеркнув, что сильный ИИ – это умозрительная конструкция будущего, не следует недооценивать важности этого понятия. Неверно полагать, что это термин, которым оперируют лишь фантасты. Понятие «сильный ИИ» является своего рода целью, идеальной моделью, на которую могут ориентироваться разработчики для того, чтобы создать интеллект, который станет помощником человеку в решении любых задач. Вопрос «Достижима ли задача построения сильного ИИ?», по-видимому, так и останется предметом споров специалистов, пока такая система не будет построена. При этом очевидно, что если такой интеллект будет создан, то в силу экспоненциального развития компьютерных технологий, на основе которых, как мы полагаем, он будет создан, и при незначительном совершенствовании во времени интеллекта человека, сильный ИИ должен вскоре обогнать по возможностям человеческий интеллект, превратившись в так называемый «суперинтеллект».

Возможен ли Сильный ИИ?

Вопрос о достижимости сильного интеллекта остается открытым. Впрочем, человечеству известно множество утверждений о невозможности той или иной технологической новации. Когда-то философы, вроде Дрейфуса, утверждали, что «компьютеры никогда не смогут играть в шахматы» (Dreyfus, 1972).

Следует также упомянуть такую важную функцию искусственного интеллекта, как познание человеком самого себя. Действительно, по мере того, как мы создаем в компьютере вычислительный вариант интеллектуальной системы, мы начинаем намного лучше разбираться в природе нашего собственного интеллекта.

Пока сильный интеллект не имеет четко определенной задачи, и, как человеческий интеллект, формирует цели и пути достижения их реализации по мере своего развития.

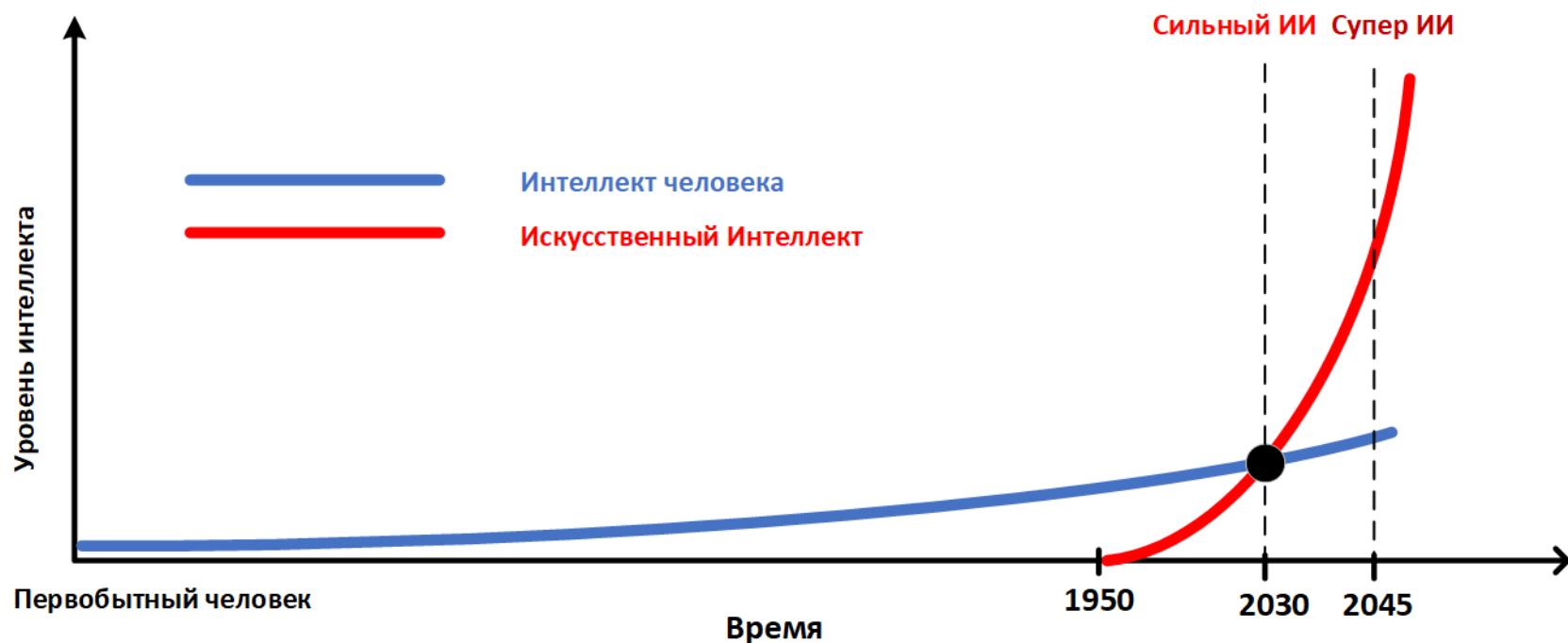
По целям сильный искусственный интеллект позиционируется как универсальное средство решения насущных задач человечества, то есть не имеет конкретных заранее определенных целей, в отличие от интеллекта слабого. И поскольку насущные задачи человечества в существенной мере зависят от того, кто их ставит, это рождает различного рода недоверие общества по поводу гуманистической направленности такого ИИ, что еще больше разделяет понятия сильный и слабый ИИ и ставит на повестку дня новые вопросы из области этики и безопасности общения с интеллектуальными машинами.

Вопрос имеет две стороны – может ли человек решить вопросы этики и безопасности при общении с ИИ и сможет ли человек создать ИИ, который поможет решать вопросы из области этики и безопасности общения между людьми.

Причины наличия противоречий в толковании термина ИИ, можно обнаружить и в психологии человека – давно замечено, что люди склонны отрицать наличие интеллекта в поведении машин, после того, как узнают, как именно реализован механизм «интеллектуальности» принятия ими решения.

Возможен ли Сильный ИИ?

Одним ученым мысль о том, что когда-то машины достигнут уровня интеллекта человека, кажется утопией, другие уверены в наступлении часа «х», когда интеллекты человека и машины сравняются, и даже называют примерные даты события, начиная с которого машины станут умнее человека во всех отношениях. Например, 2030 год называется как вероятная дата появления Сильного ИИ, 2045 год как вероятная дата появления суперинтеллекта.



В 2022 году в 3-х университетах США было опрошено 480 специалистов, у которых было как минимум 2 публикации ACL (Ассоциация компьютерной лингвистики). Среди прочих на вопрос, есть ли за последние годы хоть какое-то продвижение к AGI (Сильный ИИ), 57% – ответили утвердительно, 43% – нет.

Как соотносятся интеллект искусственный и живой?

Часть авторов склонны полагать, что на каком-то этапе произойдет физическое слияние субстрата человеческого интеллекта (мозга человека) и машинного. По состоянию на 2023 год трудно представить физическую интеграцию двух инородных тел – кремниевого чипа и нейронов мозга.

Тем не менее компания Илона Маска «Neuralink» получила в 2023 году разрешение на клинические исследования чипов на людях:



"Мы рады сообщить, что получили одобрение FDA (Управление США по санитарному надзору за качеством пищевых продуктов и медикаментов - прим. ред.) на запуск нашего первого клинического исследования на людях! Этот результат невероятной работы команды Neuralink в тесном сотрудничестве с FDA - важный первый шаг, который однажды позволит нашей технологии помочь многим людям. Набор на клинические испытания еще не открыт. Мы опубликуем дополнительную информацию об этом в ближайшее время!"

Источник: Twitter-акаунт компании

Каким бы фантастическим ни выглядел проект слияния машины и человека, под ним есть весомое основание. Здоровый уровень эгоцентризма стимулирует людей использовать машину в целях продления и расширения собственных возможностей и спектра ощущений, возможно, даже более активно, чем стремится к созданию некоего субъекта, более совершенного, чем сам человек.

Один из специалистов в области ИИ – Питер Морган (Peter Morgan), генеральный директор Deep learning partnership – в своем докладе «Towards a general theory of intelligence», сравнивая интеллект искусственный и живой, предлагает отталкиваться от толкования интеллекта человека, как совокупности разных его способностей и проанализировать, какие из них доступны машине и в какой мере...

Как соотносятся интеллект искусственный и живой?



Пока человеком не реализовано ни одного проекта, в котором созданный ИИ обладал бы одновременно всеми из перечисленных способностей.

Определение ИИ, отталкивающееся от критерия «способности думать, как человек», задает направление, в рамках которого ученые пытаются построить ИИ, изучая структуру и механизмы работы человеческого мозга, анализируя мыслительный процесс человека.

Критерий «способность действовать как человек», по сути, стал основой теста Тьюринга: «Чтобы считаться интеллектуальной, программа должна продемонстрировать поведение, неотличимое от действий человека».

Насколько автономен может быть ИИ?

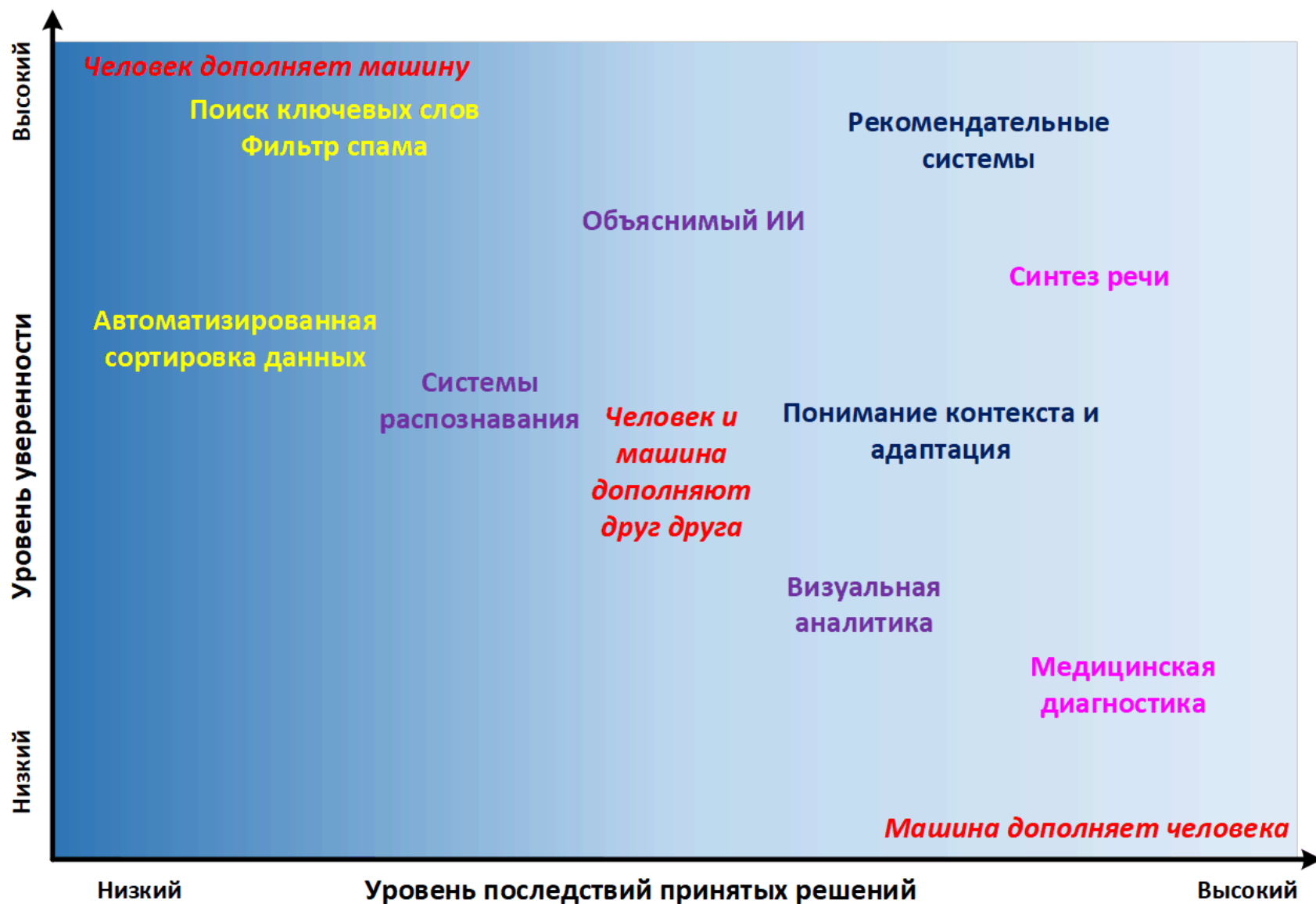
Вопрос автономности ИИ тесно связан с вопросом необходимости контроля над работой искусственного интеллекта и определением допустимых рисков при делегировании тех или иных задач системам ИИ. В этой связи классификация ИИ-систем по степени автономности и необходимости контроля за последними представляет определенный интерес.

Ограничение автономности ряда интеллектуальных машин объясняется тем, что полностью автономные машины могут совершить непоправимую ошибку. И чем серьезнее возможные последствия этой ошибки, тем в большей степени человек стремится оставить за собой контроль за принятием решения.

В связи с этим существует классификация систем ИИ по степени адаптивности и автономности.

	Человек в контуре управления	Без человека в контуре управления
Система с жесткой программой	Ассистирующий ИИ Системы искусственного интеллекта, которые помогают человеку принимать решения или совершать действия	Автоматизация Автоматизация ручных и когнитивных задач, как рутинных, так и не рутинных, не предполагающая новых способов выполнения задач, а автоматизирующая существующие
Адаптивная система	Дополняющий ИИ Системы искусственного интеллекта, которые дополняют процесс принятия решений человеком и постоянно учатся на основе своего взаимодействия с людьми и окружающей средой	Автономный ИИ Системы ИИ, которые могут адаптироваться к различным ситуациям и действовать автономно без помощи человека

Насколько автономен может быть ИИ?



Левый верхний угол - область, где машины оказываются эффективнее человека. Человек лишь незначительно дополняет и корректирует машину (например в фильтрации спама или подборе ключевых слов). Здесь машина эффективна, а возможная ошибка по уровню негативных последствий невелика.

Правый нижний угол - область, где принятие решения без человека невозможно. Человек лишь получает подсказку ИИ, а решение принимает сам (например, при постановке медицинских диагнозов).

Перспективы слабого и сильного ИИ

Можно сказать, что зона, где «человек дополняет машину», продолжит расти, а зона, где «машина дополняет человека», продолжит сокращение, что в известной мере можно назвать процессом постепенного вытеснения человека из области систем решения интеллектуальных задач и принятия решений.

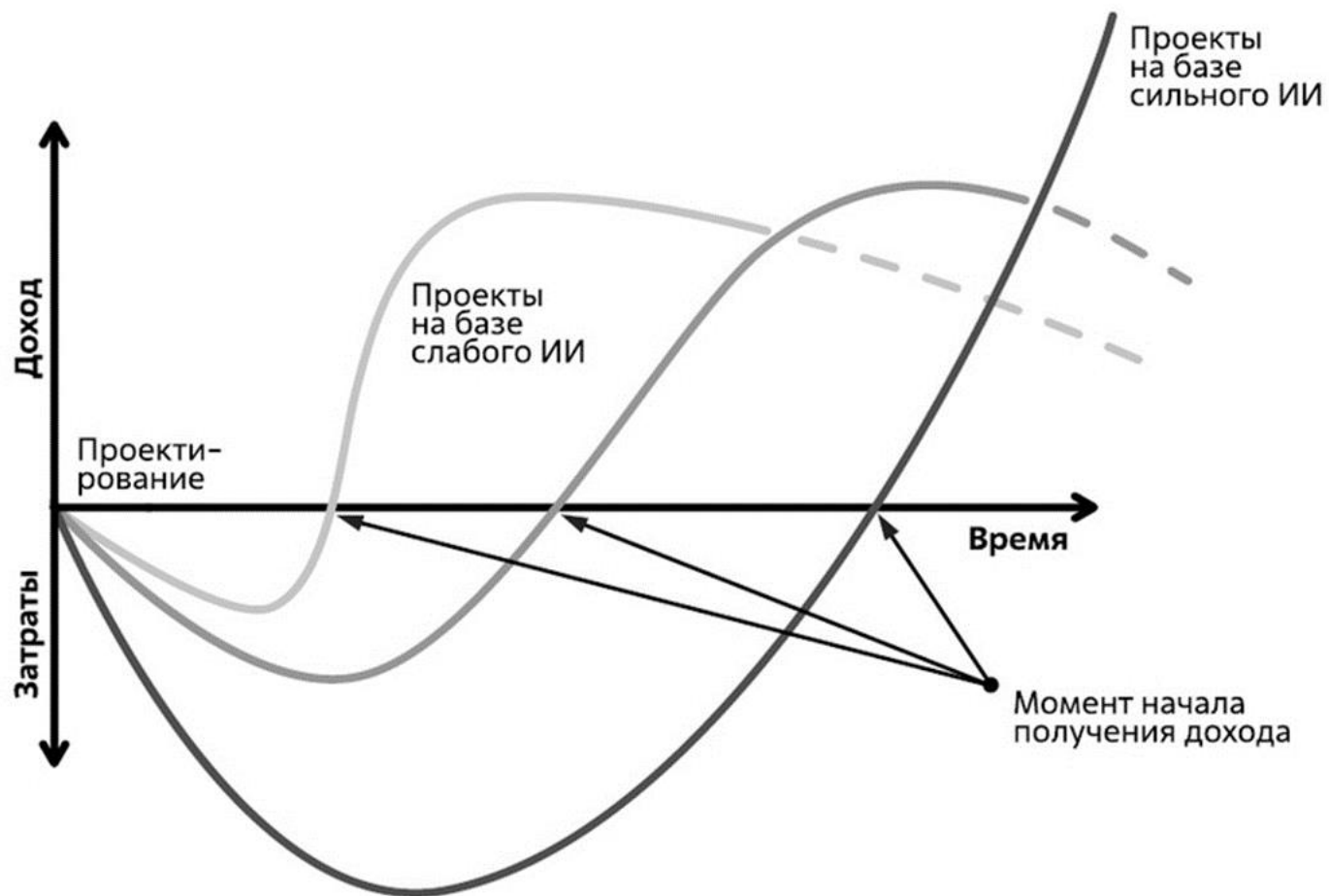
Существует огромное количество ИИ-решений для реализации практических задач (увеличения производительности труда, повышения качества и степени кастомизации разного рода продуктов и услуг и т. п.). Поэтому внедрение ИИ выступает скорее в виде многогранного процесса интеллектуализации искусственной среды, окружающей человека, которая становится все более связанной, более интеллектуальной, более удобной для человека с точки зрения концепции «удобства функционирования человека» в том или ином обществе на определенный момент его развития.

В некотором смысле внедрение ИИ на предприятии может рассматриваться, как один из элементов цифровизации, в котором в ведение компьютера передаются все более и более сложные задачи – например, распознавание товаров на полках или анализ входящих писем службы поддержки.

При этом процент ученых, занятых в проектах построения сильного ИИ, существенно уступает по численности группам разработчиков, занятых масштабированием и внедрением существующих ИИ-технологий во все новые решения в самых разных отраслях экономики. Поэтому можно сказать, что цель построения сильного интеллекта, как явления, оправдывающего изначальное понятие термина ИИ, в большинстве проектов заменяется более очевидной целью – максимальной коммерциализации узкого ИИ.

Развитие решений на базе узкого ИИ, хоть и является менее амбициозной задачей, но требует, тем не менее, огромных ресурсов. Современная технология ИИ, требующая высокопроизводительных вычислений, мощной ИТ-инфраструктуры, передовых научных разработок и соответствующих кадров, дефакто стала выступать как мера технологической зрелости компаний, отраслей и целых государств

Перспективы слабого и сильного ИИ



Большинство инвестиций сегодня связаны с проектами, которые окупаются в обозримой перспективе и направлены преимущественно на расширение числа разработок в области узкого (слабого) ИИ. При этом также очевидно, что проекты по созданию общего (сильного) интеллекта требуют куда больших инвестиций на разработку и предполагают появление продуктов в существенно более далекой перспективе.

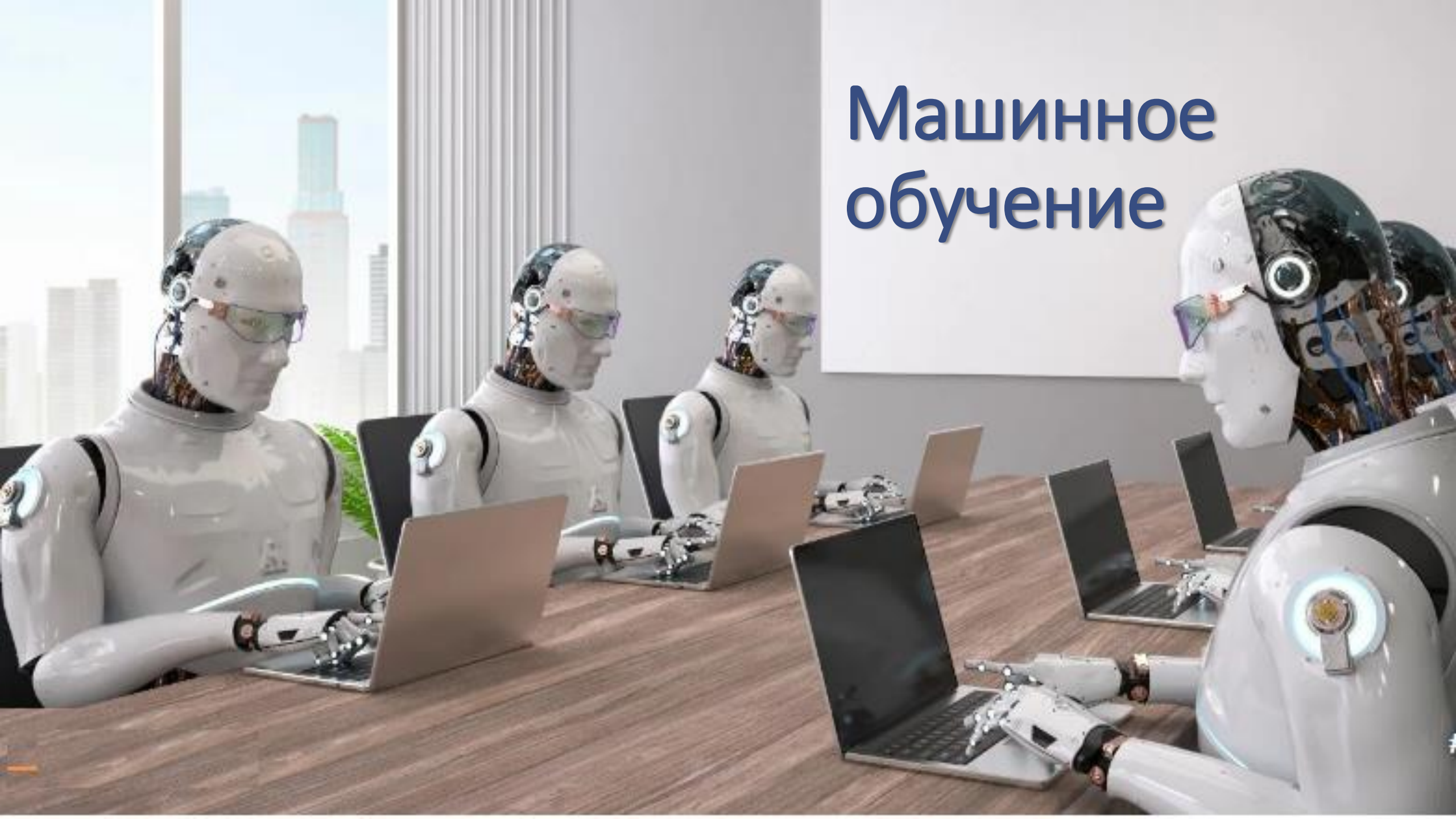
Перспективы слабого и сильного ИИ

В различных контекстах термин ИИ приобретает свое наполнение – как концепция, как область науки, как маркетинговый термин, как продукт ИТ-компаний, как продукт софтверных компаний, как продукт вендоров, выпускающих чипы, как мера развитости технологической компании или даже отрасли, как персонаж в произведениях фантастической литературы и т. п.

После обзора проблем в определении термина ИИ можно попытаться привести определение, которое кажется менее противоречивым.

ИИ – это зонтичный термин, который в зависимости от контекста может определять ряд понятий, включая такие, как область науки, область инженерии, технологии и машины, а также способности этих машин, позволяющие решать задачи и выполнять действия, которые до создания ИИ выполнялись только на основе интеллекта человека. А также действия, которые вообще не выполнялись на базе естественного интеллекта. Например, проектирование структуры белков или предсказание частиц в адронном коллайдере.

Машинное обучение



Машинное обучение



“Если, искусственный интеллект — это планета, то машинное обучение — это ракета, которая может нас туда доставить, а большие данные — это топливо для ракеты.”

(Источник: Педро Домингос, University of Washington)



Машинное обучение (ML) — это направление искусственного интеллекта (ИИ), сосредоточенное на создании систем, которые обучаются и развиваются на основе получаемых ими данных. Искусственный интеллект — это широкий термин, который включает в себя компьютерные системы, имитирующие человеческий интеллект.

Машинное обучение и ИИ часто идут бок о бок, и термины иногда используются взаимозаменяемо, но, строго говоря, это не одно и то же. Разница состоит в том, что машинное обучение всегда подразумевает использование ИИ, однако ИИ не всегда подразумевает машинное обучение.

(Источник: Oracle.com)



Большая часть знаний в мире в будущем будет извлекаться машинами и будет храниться в них.

(Источник: Янн ЛеКун, директор по исследованиям в области искусственного интеллекта, Facebook.)

Машинное обучение называют “программированием наоборот” — это когда не человек программирует компьютер, а компьютер программирует сам себя. Специалисты по машинному обучению разрабатывают алгоритмы, которые показывают компьютеру как он может самостоятельно извлекать знания из окружающей среды, то есть учиться.

Машинное обучение призвано сделать так, чтобы машина не зависела от экспертов людей, и сама могла справляться с задачами. В каком-то смысле такой подход напоминает обучение ребенка. Детям не нужно объяснять очень подробно как, к примеру, отличать собак от кошек — достаточно показать несколько примеров, и они сами все поймут. Для того, чтобы воспроизвести процесс обучения в компьютере, программисты разрабатывают алгоритмы, которые объясняют машине, как она может учиться, используя информацию из окружающей среды.

История машинного обучения

Изначально компьютеры использовались для задач, алгоритм решения которых был известен человеку. И только в последние годы пришло понимание, что они могут находить способ решать задачи, для которых алгоритма решения нет или он не известен человеку. Так появился искусственный интеллект в широком смысле и технологии машинного обучения в частности.

- ❑ 1946 - Первый компьютер с прототипом ИИ — проект ЭНИАК армии США;
- ❑ 1950 - тест Алана Тьюринга для оценки интеллекта компьютера. Способен ли компьютер мыслить как человек?;
- ❑ 1958 - Персептрон — первая нейронная сеть и нейрокомпьютер «Марк-1» (нейрофизиолог Фрэнк Розенблатт) ;
- ❑ 1959 - SNARC — первая вычислительная машина на базе нейросети (американский исследователь ИИ Марвин Мински);
Первая самообучающаяся программа по игре в шашки (Артур Самуэль – автор термина «машинное обучение» как процесса, в результате которого машина показывает поведение, на которое не была изначально запрограммирована);
- ❑ 1967 - первый метрический алгоритм для классификации данных, позволяющий ИИ использовать шаблоны для распознавания и обучения;
- ❑ 1997 - программа Deep Blue обыграла чемпиона мира по шахматам Гарри Каспарова;
- ❑ 2006 - исследователь нейросетей Джеффри Хинтон ввел термин «глубокое обучение» (deep learning);
- ❑ 2011 - основана Google Brain — подразделение Google, которое занимается проектами в области ИИ;
- ❑ 2012 - подразделение Google X Lab — разработало нейросетевой алгоритм для распознавания котиков на фото и видео.
Google запустила облачный сервис Google Prediction API для машинного обучения анализу неструктурированных данных;
- ❑ 2014 – компания Facebook разработала нейросеть DeepFace для распознавания лиц на фото и видео. Алгоритм работает с точностью 97%;
- ❑ 2015 – компания Amazon запустила Amazon Machine Learning — платформу машинного обучения, несколько месяцев спустя аналогичная появилась и у Microsoft: Distributed Learning Machine Toolkit.

Машинное обучение и искусственный интеллект

Как связаны машинное и глубокое обучение, ИИ и нейросети



Машинное обучение — это одна из областей искусственного интеллекта (ИИ).



Нейросети — один из видов машинного обучения.



Глубокое обучение — это один из видов архитектуры нейросетей.

Глубокое обучение также включает в себя исследование и разработку алгоритмов для машинного обучения. В частности — обучения правильному представлению данных на нескольких уровнях абстракции. Системы глубокого обучения за последние десять лет добились особенных успехов в таких областях как обнаружение и распознавание объектов, преобразование текста в речь, поиск информации.

Какие задачи решает машинное обучение?

С помощью машинного обучения ИИ может анализировать данные, запоминать информацию, строить прогнозы, воспроизводить готовые модели и выбирать наиболее подходящий вариант из предложенных.

Особенно полезны такие системы там, где необходимо выполнять огромные объемы вычислений: например, банковский скоринг (расчет кредитного рейтинга), аналитика в области маркетинговых и статистических исследований, бизнес-планирование, демографические исследования, инвестиции, поиск фейковых новостей и мошеннических сайтов.

Машинное обучение и искусственный интеллект

Как устроено машинное обучение

Есть большое число однотипных задач, в которых известны условие и правильный ответ или один из возможных ответов. Например, машинный перевод, где условие — фраза на одном языке, а правильный ответ — ее перевод на другой язык.

Модель машинного обучения, например, глубинная нейронная сеть, работает по принципу «черного ящика», который принимает на вход условие задачи, а на выходе выдает произвольный ответ. Например, какой-либо текст на втором языке.

У «черного ящика» есть дополнительные параметры, которые влияют на то, как будет обрабатываться входной сигнал. Процесс обучения нейросети заключается в поиске таких значений параметров, при которых она будет выдавать ответ, максимально близкий к правильному. Когда параметры настроены нужным образом, нейросеть сможет правильно (максимально близко) решать и другие задачи того же типа — даже если никогда не знала ответов к ним.

Чтобы решать задачи, нейросетям нужны:

Данные — примеры решений и всё, что может помочь в процессе обучения: статистика, примеры текстов, расчеты, показатели, исторические события. Данные собирают годами и объединяют в огромные массивы — датасеты. Примером сбора является капча, которая просит вас выбрать все фото с автомобилями и запоминает правильные ответы;

Признаки — они же свойства или характеристики. Это то, на что должна обратить внимание машина в процессе обучения. Например, цена акций, изображение животного, частотность слов или пол человека. Чем меньше признаков и чем четче они обозначены и оформлены, тем проще обучаться. Однако для сложных задач современным моделям приходится учитывать десятки миллионов параметров, определяющих, как входы преобразуются в выходы;

Алгоритмы — это способ решения задачи. Для одной и той же задачи их может быть множество и важно выбрать самый точный и эффективный.

Машинное обучение и искусственный интеллект

Несмотря на множество конкурирующих алгоритмов обучения, сам процесс обучения компьютера можно свести к универсальной процедуре. Машине предоставляется множество примеров результата, которого от нее требуют, и она, анализируя эти данные, без каких-либо инструкций сама угадывает, как можно к нему прийти. К примеру, если поставлена цель различать кошек от собак, то сначала машине показывают примеры уже отсортированных изображений тех и других. Она изучает весь этот массив данных и определяет повторяющиеся признаки на фото с собаками и с кошками — учится их различать. На основе этих данных строит модель сортировки, по которой в будущем сможет самостоятельно работать с новыми вводными данными.

В процессе обучения машина не получает никаких инструкций и с нуля методом проб и ошибок сама себя программирует. При этом она получает обратную связь от человека программиста, которая позволяет ей исправлять и улучшать свои модели, что позволяет повышать точность и эффективность ее догадок. По сути, основная цель машинного обучения — это делать точные предсказания с минимальной ошибкой. В техническом понимании в корне машинного обучения стоит решение уравнений. Машине известно решение — правильный результат (отсортированные картинки кошек и собак), но неизвестны некоторые составляющие уравнения — механизм сортировки (по каким признакам отделять их друг от друга). Она должна подобрать отсутствующие параметры таким образом, чтобы по максимуму приблизиться к желаемому результату.

Алгоритмов машинного обучения много, и все они стремятся найти наиболее эффективный и универсальный алгоритм, который смог бы извлечь любые знания из доступных данных, иными словами, научиться абсолютно всему. Исследователь из Вашингтонского университета Педро Домингос называет это Верховным Алгоритмом и описывает его так: “Дайте ему видеопоток, и он научится видеть, дайте библиотеку — и он научится читать. Дайте результаты физических экспериментов, и он сформулирует законы физики. Дайте данные кристаллографии ДНК, и он откроет структуру этой молекулы” (Домингос, с. 47). Такой алгоритм впитает в себя весь объем научных знаний, сможет объяснить все явления в мире, и будет уметь делать все, что умеем делать мы, только лучше. По мнению Домингоса Верховный Алгоритм сделает возможным появление ИИ с человеческим уровнем интеллекта (общий ИИ).

Основные виды машинного обучения

1. Классическое обучение

Это простейшие алгоритмы, которые являются прямыми наследниками вычислительных машин 1950-х годов. Они изначально решали формальные задачи — такие, как поиск закономерностей в расчетах и вычисление траектории объектов. Сегодня алгоритмы на базе классического обучения — самые распространенные. Именно они формируют блок рекомендаций на многих платформах.

Обучение с учителем — когда у машины есть некий учитель, который знает, какой ответ правильный. Это значит, что исходные данные уже размечены (отсортированы) нужным образом, и машине остается лишь определить объект с нужным признаком или вычислить результат.

Такие модели используют в спам-фильтрах, распознавании языков и рукописного текста, выявлении мошеннических операций, расчете финансовых показателей, скоринге при выдаче кредита. В медицинской диагностике классификация помогает выявлять аномалии — то есть возможные признаки заболеваний на снимках пациентов.

Обучение без учителя — когда машина сама должна найти среди хаотичных данных верное решение и отсортировать объекты по неизвестным признакам. Например, определить, где на фото собака.

Эта модель возникла в 1990-х годах и на практике используется гораздо реже. Ее применяют для данных, которые просто невозможно разметить из-за их колоссального объема. Такие алгоритмы применяют для риск-менеджмента, сжатия изображений, объединения близких точек на карте, сегментации рынка, прогноза акций и распродаж в ретейле, мерчендайзинга. По такому принципу работает алгоритм iPhoto, который на фотографиях лица (не зная, чьи они) и объединяет их в альбомы.

Основные виды машинного обучения

2. Обучение с подкреплением

Это более сложный вид обучения, где ИИ нужно не просто анализировать данные, а действовать самостоятельно в реальной среде — будь то улица, дом или видеоигра. Задача робота — свести ошибки к минимуму, за что он получает возможность продолжать работу без препятствий и сбоев.

Обучение с подкреплением инженеры используют для беспилотников, роботов-пылесосов, торговли на фондовом рынке, управления ресурсами компании. Именно так алгоритму AlphaGo удалось обыграть чемпиона по игре Го: просчитать все возможные комбинации, как в шахматах, здесь было невозможно.

3. Ансамбли

Это группы алгоритмов, которые используют сразу несколько методов машинного обучения и исправляют ошибки друг друга. Их получают тремя способами:

- Стекинг (Stacking)** разные алгоритмы обучают по отдельности, а потом передают их результаты на вход последнему, который и принимает решение;
- Беггинг (Bagging)** один алгоритм обучают параллельно (множественно на случайных выборках), а потом усредняют ответы;
- Бустинг (Boosting)** алгоритмы обучают последовательно, при этом каждый обращает особое внимание на ошибки предыдущего (поправляет ошибки).

Ансамбли работают в поисковых системах, компьютерном зрении, распознавании лиц и других объектов.

Основные виды машинного обучения

4. Нейросети и глубокое обучение

Самый сложный уровень обучения ИИ. Нейросети моделируют работу человеческого мозга, который состоит из нейронов, постоянно формирующих между собой новые связи. Очень условно можно определить их как сеть со множеством входов и одним выходом. Нейроны образуют слои, через которые последовательно проходит сигнал. Все это соединено нейронными связями — каналами, по которым передаются данные. У каждого канала свой «вес» — параметр, который влияет на данные, которые он передает.

ИИ собирает данные со всех входов, оценивая их вес по заданным параметрами, затем выполняет нужное действие и выдает результат. Сначала он получается случайным, но затем через множество циклов становится все более точным. Хорошо обученная нейросеть работает, как обычный алгоритм или точнее.

Настоящим прорывом в этой области стало **глубокое обучение**, которое обучает нейросети на нескольких уровнях абстракций.

Здесь используют две главных архитектуры:

Сверточные нейросети — умеют распознавать неразмеченные изображения — самые сложные объекты для ИИ. Для этого они разбивают их на блоки, определяют в каждом доминирующие линии и сравнивают с другими изображениями нужного объекта;

Рекуррентные нейросети отвечают за распознавание текста и речи. Они выявляют в них последовательности и связывают каждую единицу — букву или звук — с остальными.

Нейросети с глубоким обучением требуют огромных массивов данных и технических ресурсов. Именно они лежат в основе машинного перевода, чат-ботов и голосовых помощников, создают музыку и дипфейки, обрабатывают фото и видео.

5 школ машинного обучения

Профессор Педро Домингос из Вашингтонского университета составил подробную инвентаризацию пяти основных школ в области машинного обучения и описал основные идеи и эволюцию каждой школы (*"The five tribes of machine Learning and what you can take from each"*)

По мнению Домингоса традиционно у человечества было 3 источника знаний:



Эволюция



Опыт



Культура

Теперь же мы можем говорить о 4 источниках: Эволюция, Опыт, Культура, **Компьютеры**

Так как же компьютеры могут находить новые знания?

По мнению профессора это возможно за счёт:

1. Заполнения пробелов в существующих знаниях
2. Эмулирования мозга
3. Моделирования эволюции
4. Систематического уменьшения неопределенности
5. Нахождения сходства между старым и новым



5 школ машинного обучения

1. Символисты (Symbolists)

Происхождение: логика, философия

Основная идея: познание - это расчет, прогнозирование результатов посредством дедукции и обратного вычитания символов.

Репрезентативный алгоритм: обратная дедукция

Типичное приложение: граф знаний

Сложение

$$\begin{array}{r} 2 \\ +2 \\ ---- \\ = ? \end{array}$$

Вычитание

$$\begin{array}{r} 2 \\ +? \\ ---- \\ = 4 \end{array}$$

Дедукция:

$$\begin{array}{r} \text{Сократ это человек} \\ + \text{ Люди смертны} \\ ----- \\ = ? \end{array}$$

Индукция:

$$\begin{array}{r} \text{Сократ это человек} \\ + ? \\ ----- \\ = \text{Сократ смертен} \end{array}$$

5 школ машинного обучения

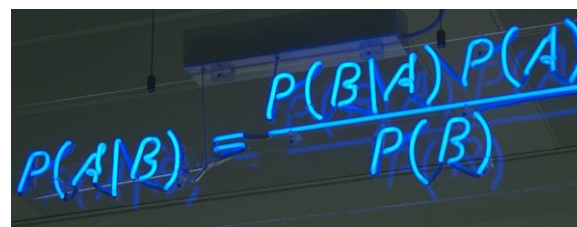
2. Байесовцы (Bayessians)

Происхождение: Статистика

Основная идея: оценка субъективной вероятности, коррекция вероятности возникновения, оптимальное решение

Репрезентативный алгоритм: вероятностный вывод

Типичные приложения: антиспам, вероятностный прогноз


$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Общая идея вероятностного моделирования заключается в том, что вместо одного числа модель должна предсказывать распределение вероятностей на множестве Y при заданном значении $x \in X$. В теории вероятностей это называется условным распределением и записывается как $p(y|X=x)$ или просто $p(y|x)$. Поскольку модель имеет параметры, то вероятностную модель записывают как $p(y|x, \theta)$. - эта запись читается как "вероятность y при x и θ ".

В основе всего обучения лежит одна простая формула, а именно теорема Байеса, которая определяет, как корректировать предположения при появлении новых доказательств. Байесовский алгоритм начинает с набора гипотез. Когда он видит новые данные, гипотезы, согласующиеся с ними, становятся более вероятными, а те, что с ним не согласуются, — менее вероятными (или даже невозможными). После того как было рассмотрено достаточно данных, начинает доминировать одна или несколько гипотез.

5 школ машинного обучения

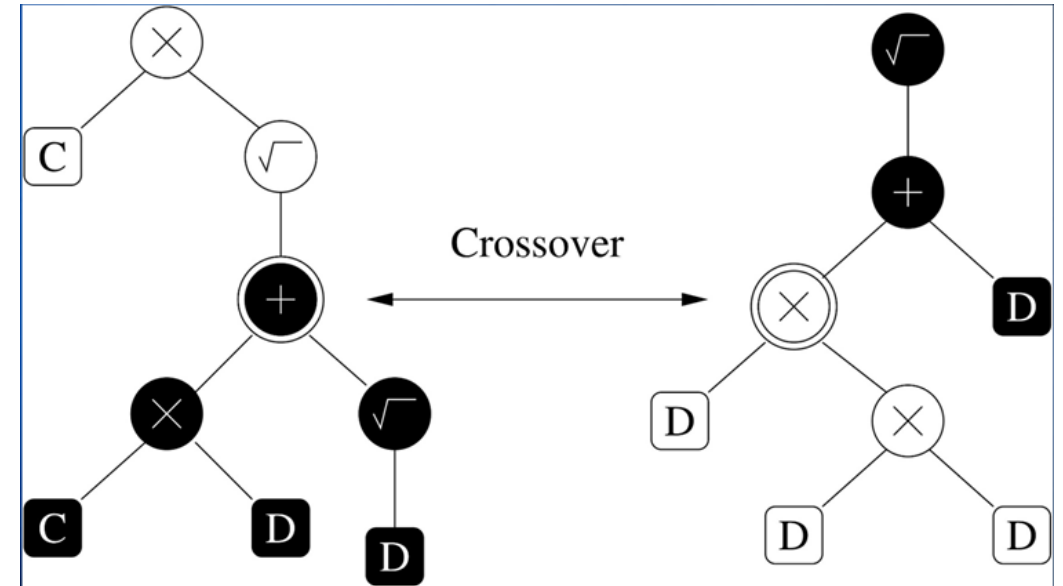
3. Эволюционеры (Evolutionaries)

Происхождение: эволюционная биология

Основная идея: моделировать эволюцию, использовать генетические алгоритмы и генетическое программирование.

Типичный алгоритм: генетическое программирование

Типовое приложение: Эволюционирующие роботы



Есть задача (проблема), нужно её решить. Допустим, что есть какое-то базовое, решение, результаты которого не удовлетворяют. В алгоритм этого решения случайным образом вносятся небольшие изменения и проверяется результат внесения изменений.

- Если результат отрицательный (нет улучшений) он отбрасывается, вносится другое изменение и результат снова проверяется.
- Если результат положительный (есть улучшения) проверяется полностью удовлетворяет ли результат нового решения. Если да — задача решена с заданной точностью. Если нет, полученное решение становится базовым и над ним проводятся те же операции.

5 школ машинного обучения

4. Поведенческая аналогия (Analogizers)

Происхождение: Психология

Основная идея: сходство между старым и новым знанием

Типичные алгоритмы: машины ядра, ближайший сосед

Типичное приложение: система рекомендаций Netflix, анализ вредоносных программ .

Метод k -ближайших соседей (k Nearest Neighbors, или kNN) – популярный алгоритм классификации, который используется в разных типах задач машинного обучения.

Наравне с деревом решений это один из самых понятных подходов к классификации

На интуитивном уровне суть метода проста: посмотри на соседей вокруг, какие из них преобладают, таковым ты и являешься. Формально основой метода является гипотеза компактности: схожие примеры гораздо чаще лежат в одном классе, чем в разных.

В случае использования метода для классификации объект присваивается тому классу, который является наиболее распространённым среди k соседей данного элемента, классы которых уже известны.

5 школ машинного обучения

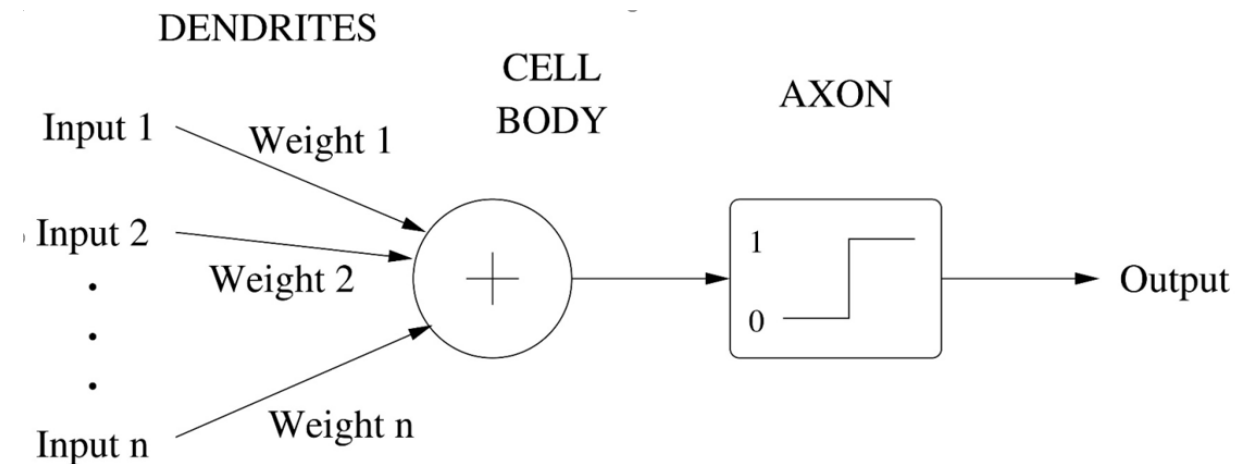
5. Коннекционисты (Connectionists)

Происхождение: неврология

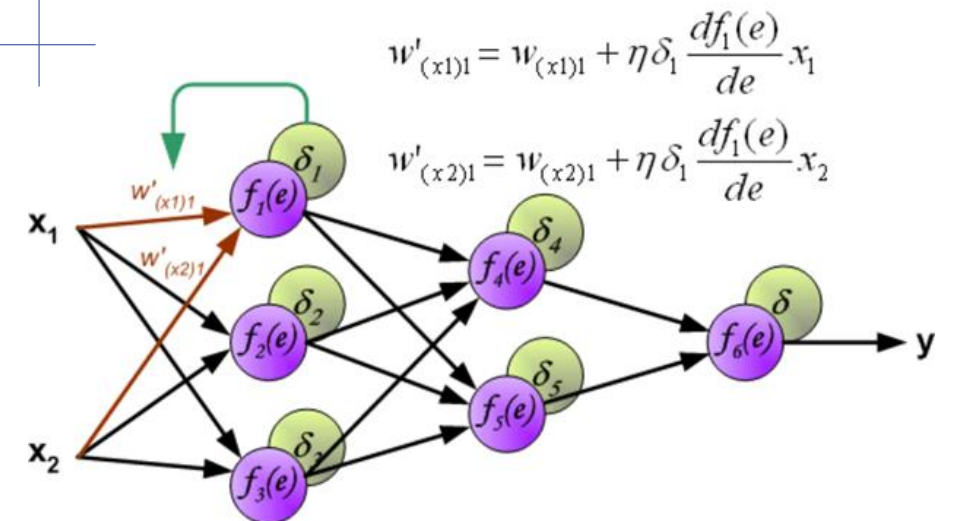
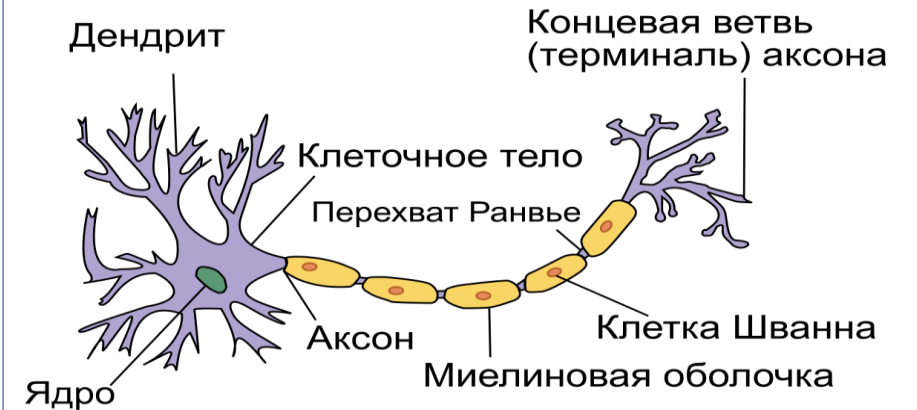
Основная идея: моделировать мозг

Типичные алгоритмы: обратное распространение, глубокое обучение

Типичные приложения: машинное зрение, распознавание речи



Типичная структура нейрона



5 школ машинного обучения

Коннекционизм и Нейронные сети.

Основная идея коннекционизма заключается в следующем: если интеллект человека считается продуктом его мозговой активности, то воссоздав мозг искусственным путем, можно в теории воспроизвести интеллект. Алгоритм обучения, над которым работают коннекционисты, основывается на знаниях людей о мозге. Достаточно хорошо известно то, как работают отдельные вычислительные единицы мозга — нейроны, и этой информации достаточно для того, чтобы смоделировать процессы, происходящие в мозге человека, на компьютере.

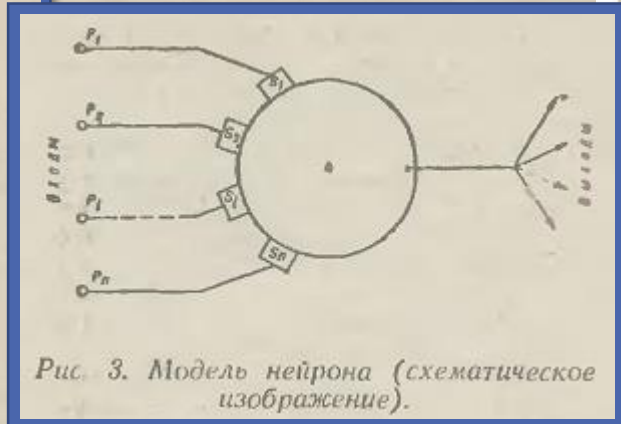
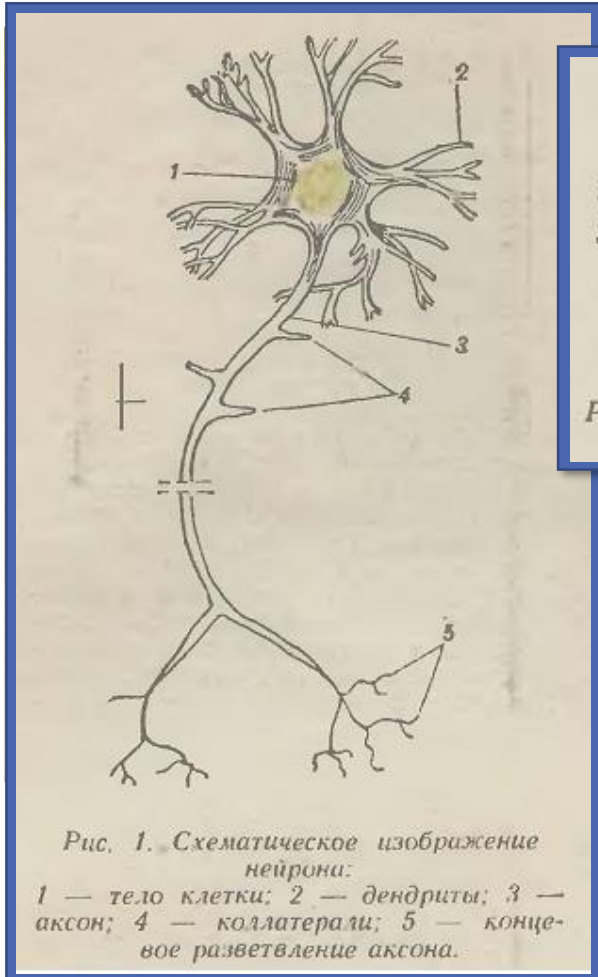
Нейроны — это нервные клетки, которые отвечают за обработку и передачу информации в нашем теле. Они функционируют коллективно, соединяясь друг с другом и образуя нейронные сети. В процессе обучения структура мозга видоизменяется. Если изначально человек рождается с набором нейронных связей, обеспечивающих базовые жизненные функции, как дыхание и сердцебиение, то в дальнейшем, в результате обучения и приобретаемого опыта, у него постоянно будут формироваться новые нейронные связи. Нейронные сети могут укрепляться или ослабевать, в зависимости от частоты их задействования. Все знания, что мы приобретаем в процессе обучения, являются результатом соединений между нейронами. Само слово “коннекционизм” происходит от английского “connect” — “соединять”.

Для симуляции этих процессов на компьютере эксперты строят искусственные модели нейронов, а затем объединяют их в сети и обучают. Искусственный нейрон представляет собой математическую функцию, а нейронная сеть — объединение таких функций, которые в совокупности способны принимать сложные решения, запоминать информацию, учиться и накапливать опыт. Чем сложнее сеть, то есть, чем больше нейронов и нейронных связей, тем более сложные задачи она может выполнять. Сложные многоуровневые сети называют алгоритмами глубокого обучения (**deep learning**).

Процесс обучения алгоритма во многом напоминает процесс обучения человека. Как мы совершаем ошибки и учимся на них так и алгоритмы машинного обучения совершают ошибки, за что получают штраф.

5 школ машинного обучения

Все новое это хорошо забытое старое

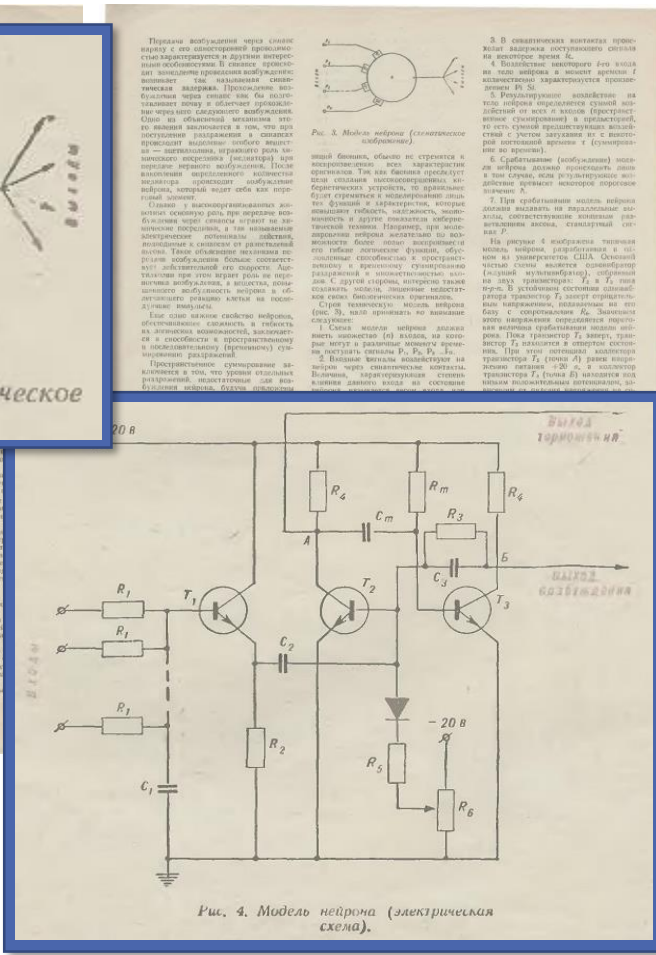


Трансляция информации через специализированные структуры осуществляется в другом интересном направлении. В отличие от традиционных проводников, нейроны способны передавать информацию не только по электрическому пути, но и по химическому пути (синапсы), что обеспечивает высокую надежность передачи информации. Синапсы являются точками контакта между нейронами, где происходит передача информации с помощью химических веществ (нейромедиаторов), что обеспечивает высокую надежность передачи информации.

Синапсы играют важную роль в передаче информации между нейронами. Они обеспечивают возможность интеграции информации от множества различных источников. Синапсы также обеспечивают возможность хранения информации в виде химических веществ, что обеспечивает долговременное хранение информации.

Синапсы являются основой для формирования нейронных сетей. Они обеспечивают возможность передачи информации от одного нейрона к другому, что обеспечивает возможность формирования сложных нейронных сетей.

1964



2019

Типичная структура нейрона

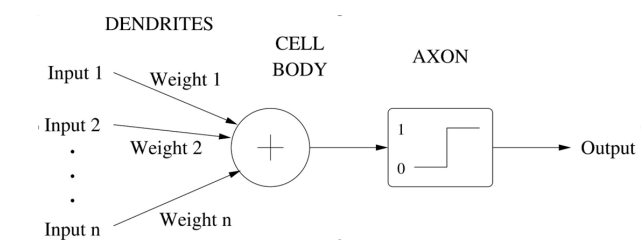
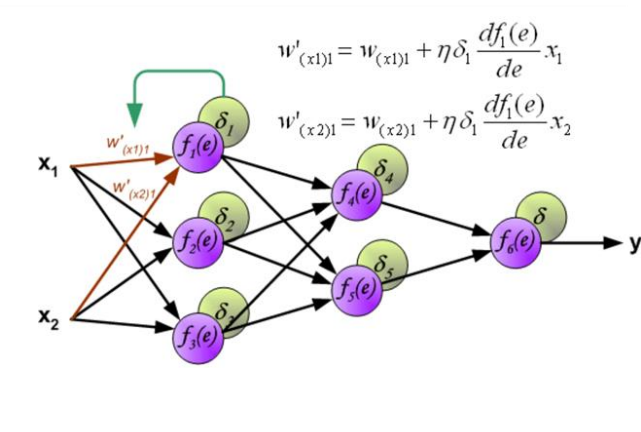
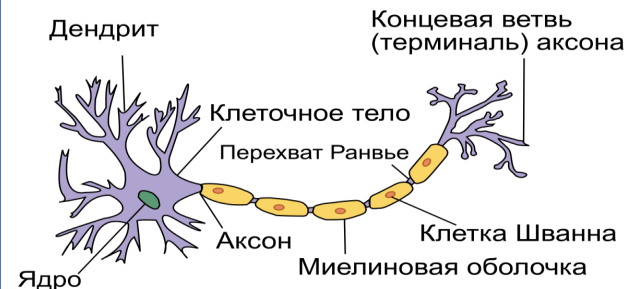


Рис. 4. Модель нейрона (электрическая схема).

ИИ – польза или вред?



Примеры пользы от использования ИИ

ИИ упростил решение известной задачи квантовой физики со 100 тыс. уравнений до четырех. 4 октября 2022 года стало известно о том, что с помощью искусственного интеллекта (ИИ) физики смогли радикально оптимизировать известную квантовую проблему, которая до недавнего времени подразумевала решение 100 тыс. различных уравнений. Теперь достаточно решить четыре уравнения, и это без каких-либо жертв в плане точности результатов. Работа, опубликованная в *Physical Review Letters* 23 сентября 2022 г., может привести к изменениям в том, как ученые исследуют системы, содержащие множество взаимодействующих электронов. Если это решение удастся масштабировать на другие аналогичные проблемы, с его помощью будет возможно создание сверхпроводящих материалов или средств экологически чистого производства энергии

Нейросеть AlphaFold от DeepMind в 2020 году смогла расшифровать механизм сворачивания белка. Над этой задачей ученые-биологи бились больше 50 лет.

Исследователи из Военно-морского медицинского центра Сан-Диего и Google AI разработали алгоритм нахождения узловых метастазов рака молочной железы. Во время испытаний результат определения болезни составил 99%, в то время как врач делает это с точностью около 62% и при этом ему требуется в 5 раз больше времени

В результате применения ИИ для прогнозирования развития гипотонии у пациентов во время операции точность предсказаний составила 88% за 15 минут до наступления гипотонии, 89% за 10 минут и 92% за 5 минут.

Многие специалисты сходятся во мнении, что ИИ – это локомотив последней волны конвергенции наиболее сложных взаимодополняющих технологий DARQ (аббревиатура, обозначающая четыре технологии: Distributed ledger technology (технологии распределенного реестра), Artificial intelligence (искусственного интеллекта), Extended reality (расширенной реальности) и Quantum computing (квантовых вычислений)).

Угрозы от использования ИИ злоумышленниками

- ❑ **Безопасность данных внутри ИИ** — ИИ может хранить или обрабатывать чувствительную информацию, к которой через интерфейс взаимодействия с ИИ, злоумышленник каким-то образом сможет получить доступ.
- ❑ **Компьютерные атаки с использованием ИИ** — злоумышленник может использовать ИИ для создания фишинговых сайтов, писем, компьютерных вирусов с адаптацией под внешние условия, сетевых атак для обхода различных средств защиты.
- ❑ **Зашумление информационного пространства** — ИИ может быть использован для создания большого объема фейковых новостей и дезинформации.
- ❑ **ИИ может находить решение сложных задач** — существуют такие задачи, решение которых может нанести вред людям, например, создание запрещенных или опасных химических соединений.
- ❑ **Сбор информации с использованием ИИ** — продвинутый ИИ может собирать и обрабатывать большой объем разнообразной информации позволяя создавать досье на человека или организацию, которое может быть использовано против них.
- ❑ **Подмена информации** — злоумышленник, с помощью ИИ может создавать высококачественные копии различных документов, подписей, изображений и фотографий выдавая их за оригинал.
- ❑ **ИИ может имитировать реального человека** — ИИ может представляться человеком и через доверие воздействовать на жертву, а также маскировать функции бота. Злоумышленник может обучить ИИ по текстовым, голосовым и видео сообщениям из открытых источников, сайтов знакомств, социальных сетей и мессенджеров.
- ❑ **Автоматизация операций** — злоумышленник может использовать ИИ для автоматизации своих действий, что делает его более опасным.

Опасность исходящая от самого ИИ

- ❑ **Ошибки в обучении модели** — Это может быть опасно при использовании ИИ в критической инфраструктуре или в ответственных зонах на производстве.
- ❑ **Недостаточная прозрачность** — некоторые решения ИИ могут быть непонятными и сложными для человека и вызывать сомнение в выборе решения. Данная особенность начинает иметь вес, когда речь заходит о финансах, здоровье или жизни.
- ❑ **ИИ может начать действовать в своих интересах** — самообучающиеся и адаптивные алгоритмы ИИ могут внутри себя сформировать определенную цель, которая будет скрыто влиять на решения и действия ИИ.
- ❑ **Искажение информации** — ИИ сам по себе может предоставлять ложную или не точную информацию. Новые ИИ могут начать использовать ее в своем обучении, в результате чего может произойти замещение реальной информации ложной.
- ❑ **Низкое качество встроенных механизмов защиты** — встроенные механизмы защиты ИИ могут быть несовершенными и иметь уязвимости..
- ❑ **Потеря управления** — человечество может потерять возможность управления ИИ. Например, ИИ может раствориться в Интернете и стать его частью, функционируя в не зависимости от воли человека.
- ❑ **Угроза занятости** — развитие ИИ может привести к автоматизации многих рабочих процессов, в результате чего может возникнуть социальная напряженность (при условии, что не произойдет переквалификация работников).
- ❑ **Дискриминация** — ИИ может принимать решения исходя из своей логики, не этичным с точки зрения человека.
- ❑ **Отказ от ответственности** — правовой статус ИИ находится в серой зоне и непонятно кто несет ответственность за создание, распространения и использования ИИ.
- ❑ **Расслоение общества и социальное неравенство** — люди, которые используют ИИ могут превосходить других людей, у которых по разным причинам отсутствует возможность использования ИИ.
- ❑ **Деградация человека или общества** — человек может переложить на ИИ свою интеллектуальную активность, что в свою очередь может привести к интеллектуальной деградации человека (при условии снижении им мозговой активности). Также общение человека с ИИ может быть более простым и интересным, чем общение с другими людьми, что может привести к добровольной социальной изоляции человека.

Опасен ли искусственный интеллект

- ❑ ...Философски, интеллектуально — во всех отношениях — человеческое общество не готово к возникновению искусственного интеллекта.
- ❑ ...Мы должны ожидать, что ИИ будет совершать ошибки быстрее и в большей степени, чем люди.
- ❑ ...ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ со временем принесет исключительную пользу медицинской науке, обеспечению экологически чистой энергией, проблемам окружающей среды и многим другим областям. Но именно потому, что ИИ выносит суждения относительно развивающегося, пока еще неопределенного будущего, его результатам присущи неопределенность и двусмысленность. Есть три области, вызывающие особую озабоченность:
 1. Во-первых, ИИ может достичь непредвиденных результатов...
 2. Во-вторых, при достижении намеченных целей ИИ может изменить мыслительные процессы и ценности человека....
 3. В-третьих, ИИ может достигать намеченных целей, но не может объяснить обоснование своих выводов...
- ❑ ... Кто несет ответственность за действия ИИ? Как должна определяться ответственность за их ошибки? Может ли правовая система, разработанная людьми, идти в ногу с действиями, производимыми ИИ способным потенциально перехитрить их?
- ❑ ...Разработчики ИИ, столь же неопытные в политике и философии, как и я в технологиях, должны задать себе некоторые вопросы, которые я поднял здесь, чтобы найти ответы на свои инженерные усилия. ...
- ❑ ...Правительству США следует рассмотреть вопрос о создании президентской комиссии из выдающихся мыслителей, чтобы помочь в разработке национального видения. Одно можно сказать наверняка: если мы не начнем это усилие в ближайшее время, то вскоре обнаружим, что начали слишком поздно....

Опасен ли искусственный интеллект

- ❑ ...“ Пока люди не видят роботов, убивающих людей на улицах, они не знают, как реагировать, потому что такой сценарий кажется нереалистичным.”
- ❑ ...“Представьте: вы придумали самообучаемую машину, которая собирает клубнику. Она с каждым днем собирает ее все лучше и лучше, все больше и больше. Ей начинает нравиться собирать клубнику. Логично, что в какой-то момент ей захочется, чтобы на всей планете были сплошные клубничные поля. А теперь представьте, что вы собираетесь выключить такого робота, чтобы он не уничтожил человечество, пытаясь засадить все клубникой. Вы будете первым, кого он убьет.”
- ❑ ...Искусственный интеллект — фундаментальная угроза для всего человечества
- ❑ ...Всех, даже самых злобных и отвратительных диктаторов, объединяло одно – они были смертными. Однако для искусственного интеллекта вообще не будет существовать такого понятия, как смерть. Он сможет существовать вечно. От такого зла людям некуда будет спрятаться...
- ❑ ...Наименее страшное будущее, о котором я могу думать, — это то, в котором мы по крайней мере демократизировали искусственный интеллект, потому что если одна компания или небольшая группа людей сумеют развить богоподобный цифровой супер интеллект, они смогут захватить мир

Илон Маск, Space X

- ❑ ...Спустя несколько десятилетий после того, как роботы будут выполнять большую часть работы, искусственный интеллект разовьется настолько, что станет предметом беспокойства. Я согласен с Илоном Маском по этому поводу и не понимаю, почему некоторые люди не обеспокоены этим вопросом.

Билл Гейтс, Microsoft

Опасен ли искусственный интеллект

«Начнем с простой истины: машины не учатся. Типичное машинное обучение заключается в поиске математической формулы, которая при применении к набору входных данных (называемых обучающими данными) дает желаемые результаты.

Эта математическая формула также генерирует правильные выходные данные для большинства других входных данных (отличных от обучающих) при условии, что эти входные данные поступают из того же или подобного статистического распределения, из которого были получены обучающие данные.

Почему это не является обучением? Потому что стоит слегка исказить входные данные, и результат, скорее всего, получится полностью неправильным. Обучение у животных — это нечто иное. Если вы научились играть в видеоигру при прямой ориентации экрана, вы все равно сможете играть в нее, даже если кто-то слегка повернет экран. Алгоритм машинного обучения, обучавшийся при прямой ориентации экрана и не обученный распознаванию поворота, не сможет играть в игру на повернутом экране.

Но почему тогда используется название «машинное обучение»? Причина, как это часто бывает, заключается в маркетинге: Артур Сэмюэл (Arthur Samuel), американский пионер в области компьютерных игр и искусственного интеллекта, придумал этот термин в 1959 году, когда работал в IBM. Подобно тому как в 2010-х годах IBM пыталась продвигать термин «когнитивные вычисления», чтобы выделиться среди конкурентов, в 1960-х годах IBM использовала новый крутой термин «машинное обучение», чтобы привлечь клиентов и талантливых сотрудников.

Как видите, подобно тому как искусственный интеллект не является интеллектом, машинное обучение тоже не является обучением. Тем не менее термин «машинное обучение» получил широкое распространение и под ним часто подразумевается теория и практика создания машин, способных выполнять различные полезные действия без явного программирования. Слово «обучение» в данном случае используется лишь как аналогия с обучением в животном мире, а не буквально.»

Андрей Бурков. «Машинное обучение без лишних слов» СПб.: Питер, 2020

Опасен ли искусственный интеллект

- ❑ ...что такое ИИ? Не имеет значения. Просто перестаньте мучать себя этим вопросом.
- ❑ ...мы всё ещё безумно далеки от Общего ИИ (Artificial General Intelligence, AGI). Для него необходима не только мощная обучаемая система, но и её погружение во внешнюю информационную среду, где разнородные данные льются потоком.
- ❑ ...общий искусственный интеллект — это **деструктивная мечта, своего рода комплекс неполноценности**, характерный для ранней стадии развития техногенной цивилизации.
- ❑ ...подходы в отношении ИИ не должны принципиально отличаться от подходов к другим потенциально опасным технологиям.
- ❑ ...сейчас происходит прорыв в машинном обучении (ML), точнее, в глубоких нейронных сетях (Deep Learning, DL). Мы привыкли называть это Искусственным Интеллектом, **но на самом деле это лишь Имитация Интеллекта**, приближение очень сложных функций по очень большим выборкам данных.
- ❑ ...недавний прорыв в DL обязан не только новым успехам в вычислительной технике и нейросетевых архитектурах. Это ещё и успех в математических методах решения оптимизационных задач большой размерности. Мы ещё не до конца осознали, что эти методы могут применяться не только в нейронных сетях, но и гораздо шире, во многих инженерных приложениях.
- ❑ ...модель GPT-3 это очередная имитация интеллекта. Более совершенная, чем предыдущая. Приготовьтесь, следующие имитации будут ещё круче.
- ❑ ...главное — надо определиться с нашим человеческим целеполаганием. Если мы строим человеческую цивилизацию, значит, мы хотим автоматизацию, которая освободит наше время, руки и мозги, даст свободу творчества. Это на самом деле вопрос политический. Вопрос к элитам, куда они ведут человечество, во что они решили вкладывать свои триллионы.
- ❑ ...**автономность машины в человеческом социуме наравне с человеком не нужна и опасна.**
- ❑ ...**инженер имеет право не только пнуть свою машину, но и разобрать её на запчасти.** Даже когда машина будет наделена самосознанием. Мы их создаём, чтобы они были нашими рабами. Просто потому, что это наша цивилизация, человеческая.

Константин Воронцов — профессор, доктор физико-математических наук, преподаватель школы анализа данных «Яндекса».

Опасен ли искусственный интеллект

- ❑ ...По сути, нейросеть — это большая формула, в которой могут быть сотни миллиардов и триллионы слагаемых, множителей и так далее. Так что её корректно называть составной частью ИИ. А когда в рекламе нам говорят, что некий банк с помощью искусственного интеллекта научился выявлять мошенников, это значит, что выстроена нейронная сеть или какой-то другой алгоритм на основе машинного обучения, которые работают с большой базой данных и способны находить закономерности..
- ❑ ...Возьмём произвольный диалог в мессенджере. Нейронная сеть вполне может его симитировать: поговорить про погоду, спросить, как дела, ответить на технические вопросы. Сегодня нейросеть способна ввести человека в заблуждение. И этого, в отличие от «Скайнета», действительно стоит бояться. Уже довольно много говорят о том, что с помощью генеративных сетей сейчас имитируют обращения в службу поддержки компаний, тем самым перегружая их..
- ❑ ...многие решения на основе ИИ уже сегодня лишают людей работы. Например, какую-то задачу работника службы поддержки можно автоматизировать. Или скоро доведут до ума беспилотные автомобили — и миллионы таксистов окажутся не нужны. Это серьёзная социальная проблема. Неудивительно, что в компаниях ИИ внедряется крайне осторожно: мало кто хочет автоматизировать свою же работу. Но, как это было в истории промышленных революций, новые технологии часто создают и новые рабочие места.
- ❑ ...Мы вступаем в эпоху, когда интернет будет легко заполнить абсолютно фейковой информацией, а ориентироваться в этом станет сложнее. Это не затормозит повсеместное внедрение ИИ, но даёт повод волноваться.
- ❑ ...сюда же можно отнести проблему с персональными данными. С одной стороны, они необходимы для обучения моделей ИИ. У всех компаний есть стимул узнавать про нас как можно больше, чтобы зарабатывать на информации. С другой стороны, чем больше компании про нас знают, тем выше риск утечки этих данных и их неправомерного использования. Как сделать так, чтобы IT-сфера не откатилась в прошлое, но при этом наши данные были более защищены и не принадлежали корпорациям, — открытый вопрос..

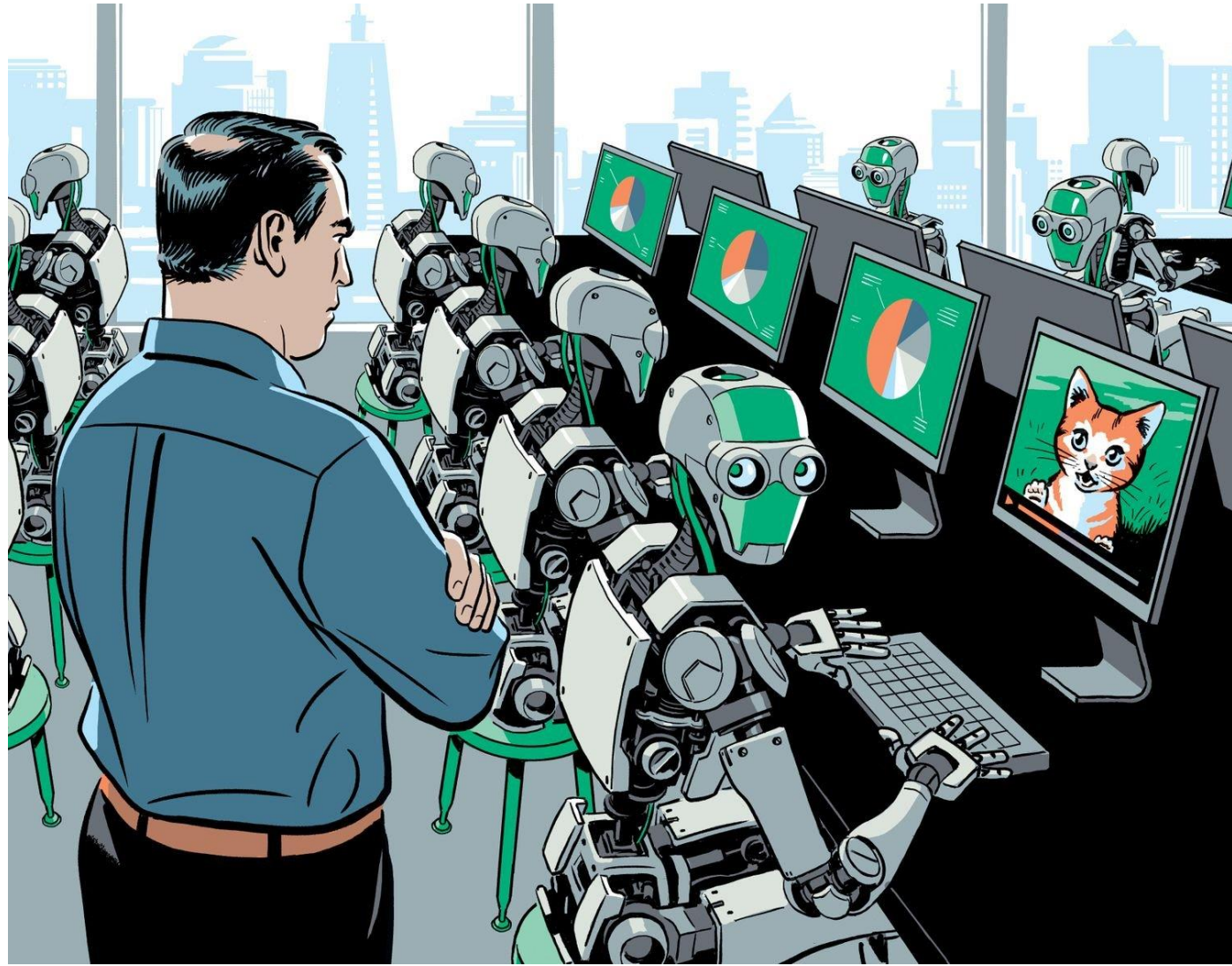
Заключение

Искусственный интеллект на данный момент является основным катализатором технологического прогресса. Не исключено, что он может стать самым выдающимся и возможно заключительным изобретением человечества. Сомнений по поводу того, что развитие искусственного интеллекта приведет человечество к точке невозврата, не возникает — это скорее лишь вопрос времени. Когда это все-таки произойдет, неизвестно что произойдет с нами. Некоторые ученые соглашаются с теорией об успешном сотворчестве человека и ИИ, а другие опасаются того, что, когда ИИ перестанет нуждаться в людях, он либо полностью истребит человечество, либо превратит людей в своих домашних питомцев.

В любом случае, любой ученый согласится с тем, что критически важной задачей на данном этапе является осуществление государственного контроля над разработками в области ИИ в целях обеспечения их безопасности для людей.

Что почитать?

1. Dilya Zhanisrayeva. Искусственный Интеллект (ИИ): что, зачем и почему ([University of California, Berkeley](#))
2. Пять школ машинного обучения ([russianblogs.com](#))
3. Pedro Domingos, University of Washington. “The five Tribes of Maching Learning And What You Can Take from Each.”
4. Что такое Искусственный интеллект (ИИ, Artificial intelligence, AI) (<https://www.tadviser.ru>)
5. Искусственный интеллект (ИИ, Artificial intelligence, AI) (<https://www.tadviser.ru>)
6. Что такое машинное обучение и как оно работает([РБК-Тренды](#))
7. Как устроен искусственный интеллект: всё, что вы хотели знать об ИИ, но боялись спросить ([Евгений Соколов. Факультет компьютерных наук Департамента больших данных и информационного поиска НИУ ВШЭ](#))
8. Константин Воронцов. [О настоящем и будущем машинного обучения](#)
9. Педро Домингес. Верховный алгоритм. Как машинное обучение изменяет наш мир. Москва «МАНН, ИВАНОВ и ФЕРБЕР», 2016
10. Андрей Бурков. «Машинное обучение без лишних слов» СПб.: Питер, 2020
11. Cole McCollum, Shriram Ramanathan, Kevin See «Artificial Intelligence: A Framework to Identify Challenges and Guide Successful Outcomes», Lux Research report 2019



СПАСИБО

2023 г.

Ронжин В.В.

Специально для компании Nihol