



Искусственный интеллект

Искусственный интеллект (**ИИ**) — это способность вычислительных систем выполнять задачи, обычно связанные с человеческим интеллектом, такие как обучение, рассуждение, решение проблем, восприятие и принятие решений. Это область исследований в области компьютерных наук, которая разрабатывает и изучает методы и программное обеспечение, позволяющие машинам воспринимать окружающую среду и использовать обучение и интеллект для принятия мер, максимально повышающих их шансы на достижение поставленных целей. ^[1]

К числу наиболее известных приложений ИИ относятся продвинутые поисковые системы (например, Google Search); рекомендательные системы (используемые YouTube, Amazon и Netflix); виртуальные помощники (например, Google Assistant, Siri и Alexa); автономные транспортные средства (например, Waymo); генеративные и творческие инструменты (например, языковые модели и искусство ИИ); а также сверхчеловеческая игра и анализ в стратегических играх (например, шахматы и го). Однако многие приложения ИИ не воспринимаются как ИИ: «Многие передовые ИИ просочились в общие приложения, часто не называясь ИИ, потому что как только что-то становится достаточно полезным и достаточно распространенным, оно больше не называется ИИ». ^{[2][3]}

Различные подобласти исследований ИИ сосредоточены вокруг конкретных целей и использования конкретных инструментов. Традиционные цели исследований ИИ включают обучение, рассуждение, представление знаний, планирование, обработку естественного языка, восприятие и поддержку робототехники. ^[a] Для достижения этих целей исследователи ИИ адаптировали и интегрировали широкий спектр методов, включая поиск и математическую оптимизацию, формальную логику, искусственные нейронные сети и методы, основанные на статистике, исследовании операций и экономике. ^[b] ИИ также опирается на психологию, лингвистику, философию, нейронауку и другие области. ^[4] Некоторые компании, такие как OpenAI, Google DeepMind и Meta, ^[5] стремятся создать общий искусственный интеллект (AGI) — ИИ, который может выполнить практически любую когнитивную задачу, по крайней мере, так же хорошо, как человек.

Искусственный интеллект был основан как академическая дисциплина в 1956 году, ^[6] и эта область прошла через несколько циклов оптимизма на протяжении своей истории, ^{[7][8]} за которыми следовали периоды разочарования и потери финансирования, известные как зимы ИИ. ^{[9][10]} Финансирование и интерес значительно возросли после 2012 года, когда графические процессоры начали использоваться для ускорения нейронных сетей, а глубокое обучение превзошло предыдущие методы ИИ. ^[11] Этот рост еще больше ускорился после 2017 года с архитектурой трансформатора. ^[12] В 2020-х годах продолжающийся период быстрого прогресса в передовом генеративном ИИ стал известен как бум ИИ. Способность генеративного ИИ создавать и изменять контент привела к нескольким непреднамеренным последствиям и вреду, что вызвало этические опасения по поводу долгосрочных эффектов ИИ и потенциальных экзистенциальных рисков, вызвав дискуссии о регуляторной политике для обеспечения безопасности и преимуществ технологии.

Цели

Общая проблема моделирования (или создания) интеллекта была разбита на подзадачи. Они состоят из конкретных характеристик или возможностей, которые исследователи ожидают от интеллектуальной системы. Описанные ниже характеристики привлекли наибольшее внимание и охватывают область исследований в области искусственного интеллекта. ^[a]

Рассуждение и решение проблем

Ранние исследователи разработали алгоритмы, имитирующие пошаговые рассуждения, которые люди используют при решении головоломок или при выполнении логических выводов. ^[13] К концу 1980-х и 1990-х годов были разработаны методы работы с неточной или неполной информацией, использующие концепции теории вероятностей и экономики. ^[14]

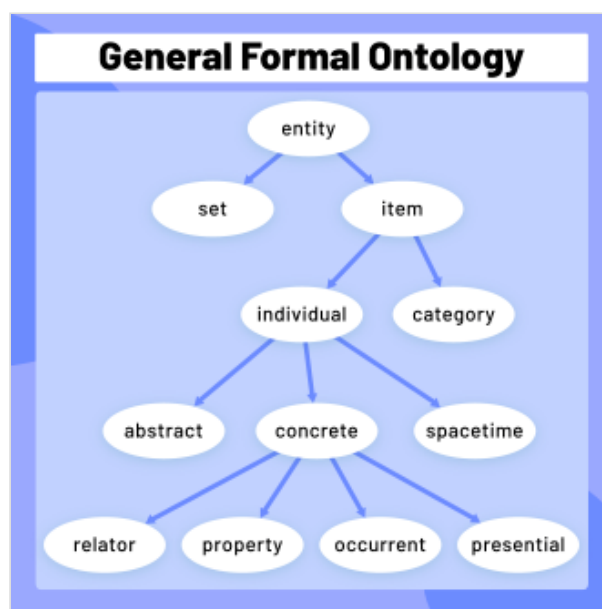
Многие из этих алгоритмов недостаточны для решения сложных задач рассуждения, поскольку они испытывают «комбинаторный взрыв»: они становятся экспоненциально медленнее по мере роста задач. ^[15] Даже люди редко используют пошаговый вывод, который могли моделировать ранние исследования ИИ. Они решают большинство своих задач, используя быстрые, интуитивные суждения. ^[16] Точное и эффективное рассуждение — нерешённая проблема.

Представление знаний

Представление знаний и инженерия знаний ^[17] позволяют программам искусственного интеллекта разумно отвечать на вопросы и делать выводы о фактах реального мира. Формальные представления знаний используются для индексации и поиска информации на основе контента, ^[18] интерпретации сцен, ^[19] поддержки принятия клинических решений, ^[20] обнаружения знаний (извлечения «интересных» и практических выводов из больших баз данных) ^[21] и других областях. ^[22]

База знаний — это совокупность знаний, представленная в форме, пригодной для использования программой. Онтология — это набор объектов, отношений, концепций и свойств, используемых в определённой области знаний. ^[23] Базы знаний должны представлять

такие вещи, как объекты, свойства, категории и отношения между объектами; ^[24] ситуации, события, состояния и время; ^[25] причины и следствия; ^[26] знания о знаниях



Онтология представляет знания как набор концепций в рамках предметной области и взаимосвязей между этими концепциями.

(то, что мы знаем о том, что знают другие люди); ^[27] рассуждения по умолчанию (то, что люди считают истинным, пока им не скажут иначе, и что это останется истинным даже при изменении других фактов); ^[28] и многие другие аспекты и области знаний.

Среди наиболее сложных проблем представления знаний – широта повседневных знаний (набор элементарных фактов, известных среднему человеку, огромен); ^[29] и субсимвольная форма большинства повседневных знаний (многое из того, что знают люди, не представлено в виде «фактов» или «утверждений», которые они могли бы выразить вербально). ^[16] Существует также сложность приобретения знаний, проблема получения знаний для приложений ИИ. ^[c]

Планирование и принятие решений

«Агент» — это всё, что воспринимает окружающий мир и совершает в нём действия. Рациональный агент имеет цели или предпочтения и предпринимает действия для их достижения. ^[d] ^[32] При автоматизированном планировании у агента есть конкретная цель. ^[33] При автоматизированном принятии решений у агента есть предпочтения: есть ситуации, в которых он предпочёл бы оказаться, и ситуации, которых он старается избегать. Агент, принимающий решения, присваивает каждой ситуации число (называемое «полезностью»), которое измеряет степень её предпочтительности для агента. Для каждого возможного действия он может рассчитать «ожидаемую полезность»: полезность всех возможных результатов действия, взвешенную по вероятности его наступления. Затем он может выбрать действие с максимальной ожидаемой полезностью. ^[34]

В классическом планировании агент точно знает, каким будет эффект любого действия. ^[35] Однако в большинстве реальных задач агент может не быть уверен в ситуации, в которой он находится (она «неизвестна» или «ненаблюдаема»), и он может не знать наверняка, что произойдёт после каждого возможного действия (она не «детерминирована»). Он должен выбрать действие, сделав вероятностное предположение, а затем переоценить ситуацию, чтобы убедиться, сработало ли действие. ^[36]

В некоторых задачах предпочтения агента могут быть неопределёнными, особенно если в них участвуют другие агенты или люди. Их можно изучить (например, с помощью обратного обучения с подкреплением), или агент может искать информацию для улучшения своих предпочтений. ^[37] Теория ценности информации может использоваться для оценки ценности исследовательских или экспериментальных действий. ^[38] Пространство возможных будущих действий и ситуаций, как правило, чрезвычайно велико, поэтому агенты должны предпринимать действия и оценивать ситуации, будучи неуверенными в результате.

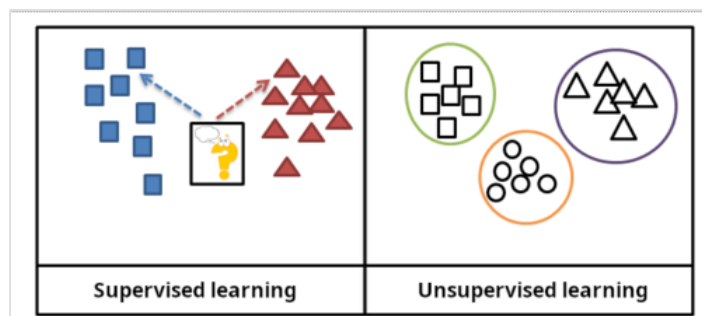
Марковский процесс принятия решений включает в себя модель перехода, описывающую вероятность того, что конкретное действие изменит состояние определённым образом, и функцию вознаграждения, которая определяет полезность каждого состояния и стоимость каждого действия. Политика связывает решение с каждым возможным состоянием. Политика может быть рассчитана (например, путём итерации), быть эвристической или может быть обучена. ^[39]

Теория игр описывает рациональное поведение множества взаимодействующих агентов и используется в программах искусственного интеллекта, которые принимают решения, в которых участвуют другие агенты. ^[40]

Обучение

Машинное обучение — это изучение программ, которые могут автоматически повышать свою производительность при выполнении заданной задачи. ^[41] Оно было частью искусственного интеллекта с самого начала. ^[e]

Существует несколько видов машинного обучения. Неконтролируемое обучение анализирует поток данных, находит закономерности и делает прогнозы без какого-либо дополнительного руководства. ^[44] Контролируемое обучение требует маркировки обучающих данных ожидаемыми ответами и бывает двух основных видов: классификация (где программа должна научиться предсказывать, к какой категории относятся входные данные) и регрессия (где программа должна выводить числовую функцию на основе числовых входных данных). ^[45]



При контролируемом обучении обучающие данные маркируются ожидаемыми ответами, тогда как при неконтролируемом обучении модель выявляет закономерности или структуры в немаркированных данных.

При обучении с подкреплением агент получает вознаграждение за хорошие ответы и наказание за плохие. Агент учится выбирать ответы, которые классифицируются как «хорошие». ^[46] Перенос обучения — это применение знаний, полученных при решении одной проблемы, к новой. ^[47] Глубокое обучение — это тип машинного обучения, при котором входные данные обрабатываются через биологически вдохновлённые искусственные нейронные сети для всех этих типов обучения. ^[48]

Теория вычислительного обучения может оценивать учащихся по вычислительной сложности, по сложности выборки (сколько требуется данных) или по другим понятиям оптимизации. ^[49]

Обработка естественного языка

Обработка естественного языка (NLP) позволяет программам читать, писать и общаться на человеческих языках. ^[50] К конкретным проблемам относятся распознавание речи, синтез речи, машинный перевод, извлечение информации, поиск информации и ответы на вопросы. ^[51]

Ранние работы, основанные на генеративной грамматике и семантических сетях Ноама Хомского, испытывали трудности с разрешением смысловой неоднозначности слов ^[f], если только не ограничивались небольшими областями, называемыми «микромирами» (из-за проблемы знания здравого смысла ^[29]). Маргарет Мастерман считала, что ключом к пониманию языков является значение, а не грамматика, и что основой вычислительной языковой структуры должны быть тезаурусы, а не словари.

Современные методы глубокого обучения для обработки естественного языка включают в себя встраивание слов (представление слов, как правило, в виде векторов, кодирующих их значение), ^[52] трансформаторы (архитектура глубокого обучения, использующая механизм

внимания), [53] и другие. [54] В 2019 году генеративные предварительно обученные языковые модели трансформаторов (или «GPT») начали генерировать связный текст, [55][56] и к 2023 году эти модели смогли получить баллы человеческого уровня на экзамене на адвоката , тесте SAT , тесте GRE и многих других реальных приложениях. [57]

Восприятие

Машинное восприятие — это способность использовать данные с датчиков (таких как камеры, микрофоны, беспроводные сигналы, активные лидары , сонары, радары и тактильные датчики) для оценки различных аспектов окружающего мира. Компьютерное зрение — это способность анализировать визуальные данные. [58]

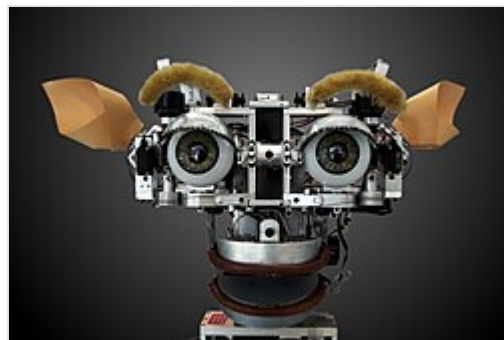
Область включает распознавание речи , [59] классификацию изображений , [60] распознавание лиц , распознавание объектов , [61] отслеживание объектов , [62] и роботизированное восприятие . [63]

Социальный интеллект

Аффективные вычисления — это область, которая включает в себя системы, которые распознают, интерпретируют, обрабатывают или имитируют человеческие чувства, эмоции и настроение . [65] Например, некоторые виртуальные помощники запрограммированы говорить разговорным языком или даже шутить; это делает их более чувствительными к эмоциональной динамике человеческого взаимодействия или иным образом облегчает взаимодействие человека с компьютером .

Однако это имеет тенденцию давать наивным пользователям нереалистичное представление об интеллекте существующих компьютерных агентов. [66]

Умеренные успехи, связанные с аффективными вычислениями, включают текстовый анализ настроений и, в последнее время, мультимодальный анализ настроений , в котором ИИ классифицирует эффекты, демонстрируемые субъектом, записанным на видео. [67]



Kismet , роботизированная голова, созданная в 1990-х годах; это машина, которая может распознавать и имитировать эмоции. [64]

Общий интеллект

Машина с искусственным интеллектом общего назначения сможет решать широкий спектр задач с широтой и универсальностью, сравнимой с человеческим интеллектом . [68]

Методы

Исследования в области искусственного интеллекта используют широкий спектр методов для достижения вышеуказанных целей. [b]

Поиск и оптимизация

ИИ может решать множество проблем, осуществляя интеллектуальный поиск среди множества возможных решений. ^[69] В ИИ используются два совершенно разных вида поиска: поиск в пространстве состояний и локальный поиск .

Поиск государственного пространства

Поиск в пространстве состояний просматривает дерево возможных состояний, чтобы попытаться найти целевое состояние. ^[70] Например, алгоритмы планирования просматривают деревья целей и подцелей, пытаясь найти путь к заданной цели, процесс, называемый анализом средств и целей . ^[71]

Простые исчерпывающие поиски ^[72] редко бывают достаточными для решения большинства реальных задач: пространство поиска (количество мест для поиска) быстро разрастается до астрономических значений . В результате поиск становится слишком медленным или вообще не завершается. ^[15] « Эвристика » или « практические правила » могут помочь расставить приоритеты в выборе вариантов, которые с большей вероятностью приведут к достижению цели. ^[73]

Состязательный поиск используется в игровых программах, таких как шахматы или го. Он просматривает дерево возможных ходов и контрходов в поисках выигрышной позиции. ^[74]

Локальный поиск

Локальный поиск использует математическую оптимизацию для поиска решения задачи. Он начинается с некоторой догадки и постепенно её уточняет. ^[75]

Градиентный спуск — это тип локального поиска, который оптимизирует набор числовых параметров путём их постепенной корректировки для минимизации функции потерь . Различные варианты градиентного спуска обычно используются для обучения нейронных сетей ^[76] с помощью алгоритма обратного распространения .

Другой тип локального поиска — эволюционные вычисления , целью которых является итеративное улучшение набора возможных решений путем их «мутации» и «рекомбинации», выбирая только наиболее приспособленных для выживания в каждом поколении. ^[77]

Распределённые процессы поиска могут координироваться с помощью алгоритмов роевого интеллекта . Два популярных алгоритма роевого интеллекта, используемых в поиске, — это оптимизация роя частиц (вдохновлённая птичьими стаями) и оптимизация колонии муравьёв (вдохновлённая муравьиными тропами). ^[78]

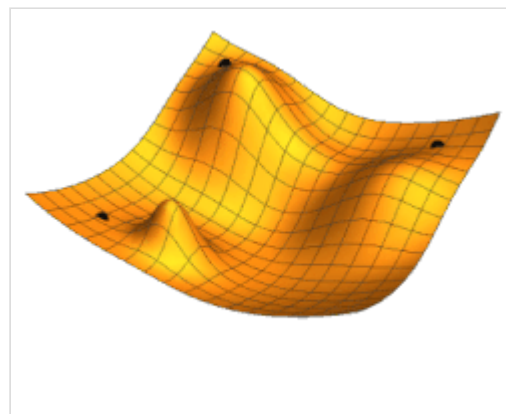


Иллюстрация градиентного спуска для трёх различных начальных точек; два параметра (представленные координатами плана) корректируются для минимизации функции потерь (высоты)

Логика

Формальная логика используется для рассуждений и представления знаний . [79] Формальная логика существует в двух основных формах: пропозициональная логика (которая оперирует утверждениями, которые являются истинными или ложными, и использует логические связки , такие как «и», «или», «не» и «подразумевает») [80] и предикатная логика (которая также оперирует объектами, предикатами и отношениями и использует квантификаторы , такие как «*Каждый X есть Y* » и «*Существуют некоторые X , которые являются Y* »). [81]

Дедуктивное рассуждение в логике — это процесс доказательства нового утверждения (вывода) на основе других утверждений, которые даны и считаются истинными (предпосылки). [82] Доказательства могут быть структурированы в виде деревьев доказательств , в которых узлы помечены предложениями, а дочерние узлы соединены с родительскими узлами правилами вывода .

При наличии проблемы и набора предпосылок решение проблемы сводится к поиску дерева доказательств, корневой узел которого помечен решением проблемы, а конечные узлы помечены предпосылками или аксиомами . В случае предложений Хорна поиск решения проблемы может быть выполнен путем рассуждения вперед от предпосылок или назад от проблемы. [83] В более общем случае клаузуальной формы логики первого порядка разрешение представляет собой единое правило вывода, свободное от аксиом, в котором проблема решается путем доказательства противоречия из предпосылок, которые включают отрицание решаемой проблемы. [84]

Вывод как в логике хорновских предложений, так и в логике первого порядка неразрешим , а следовательно, и не поддаётся обработке . Однако обратный вывод с хорновскими предложениями, лежащий в основе вычислений в языке логического программирования Пролог , является полным по Тьюрингу . Более того, его эффективность сопоставима с вычислением в других символических языках программирования. [85]

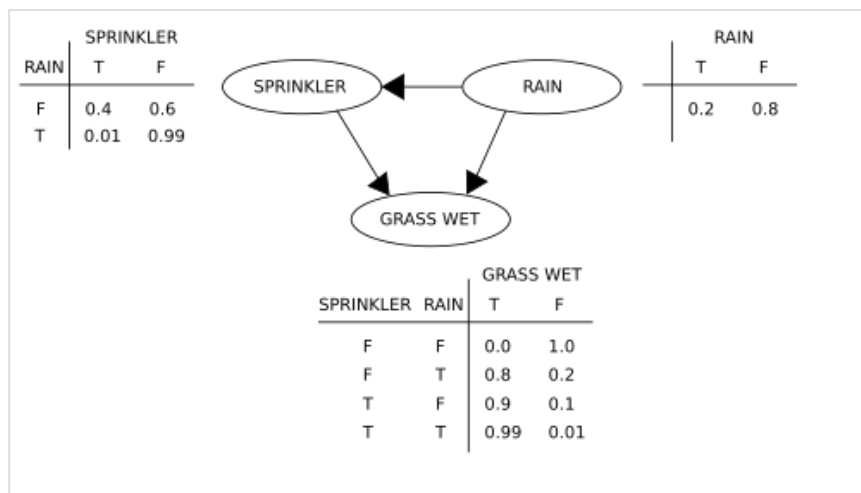
Нечеткая логика присваивает «степень истинности» от 0 до 1. Поэтому она может обрабатывать предложения, которые являются неопределенными и частично истинными. [86]

Немонотонная логика , включая логическое программирование, в котором отрицание рассматривается как неудача , предназначена для обработки рассуждений по умолчанию . [28] Были разработаны и другие специализированные версии логики для описания многих сложных областей.

Вероятностные методы для неопределенных рассуждений

Многие проблемы в ИИ (включая рассуждение, планирование, обучение, восприятие и робототехнику) требуют, чтобы агент работал с неполной или неточной информацией. Исследователи ИИ разработали ряд инструментов для решения этих проблем, используя методы теории вероятностей и экономики. [87] Были разработаны точные математические инструменты, которые анализируют, как агент может делать выбор и планировать, используя теорию принятия решений , анализ решений [88] и теорию информационной

ценности



Простая байесовская сеть с соответствующими таблицами условных вероятностей

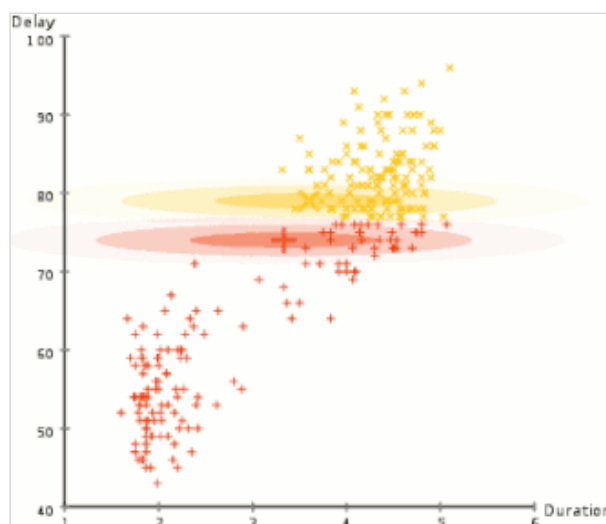
[89] Эти инструменты включают такие модели , как марковские процессы принятия решений , [90] динамические сети принятия решений , [91] теория игр и проектирование механизмов . [92]

Байесовские сети [93] – это инструмент, который можно использовать для рассуждений (с использованием алгоритма байесовского вывода), [94] [95] обучения (с использованием алгоритма максимизации ожидания), [96] [97] планирования (с использованием сетей принятия решений) [98] и восприятия (с использованием динамических байесовских сетей). [91]

Вероятностные алгоритмы также могут использоваться для фильтрации, прогнозирования, сглаживания и поиска объяснений для потоков данных, помогая таким образом системам восприятия анализировать процессы, происходящие с течением времени (например, скрытые модели Маркова или фильтры Калмана). [91]

Классификаторы и статистические методы обучения

Простейшие приложения ИИ можно разделить на два типа: классификаторы (например, «если блестит, то бриллиант»), с одной стороны, и контроллеры (например, «если бриллиант, то подбери»), с другой стороны. Классификаторы [99] — это функции, которые используют сопоставление с образцом для определения наиболее близкого соответствия. Их можно точно настроить на основе выбранных примеров с помощью контролируемого обучения . Каждый образ (также называемый «наблюдением ») помечен определенным предопределенным классом. Все наблюдения вместе с их метками классов называются набором данных . При получении нового наблюдения оно классифицируется на основе предыдущего опыта. [45]



Ожидание-максимизация кластеризации данных об извержениях вулкана Старый Служак начинается со случайного предположения, но затем успешно сходится к точной кластеризации двух физически различных режимов извержения.

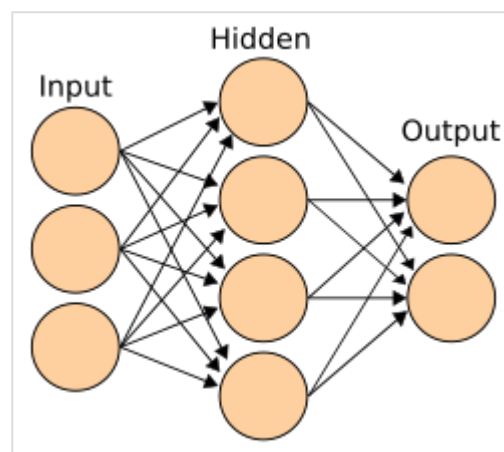
Существует множество видов классификаторов. ^[100] Дерево решений — самый простой и наиболее широко используемый символический алгоритм машинного обучения. ^[101] Алгоритм k-ближайших соседей был наиболее широко используемым аналоговым ИИ до середины 1990-х годов, а методы ядра, такие как машина опорных векторов (SVM), вытеснили k-ближайших соседей в 1990-х годах. ^[102] Сообщается, что наивный байесовский классификатор является «наиболее широко используемым обучающимся алгоритмом» ^[103] в Google, отчасти благодаря своей масштабируемости. ^[104] Нейронные сети также используются в качестве классификаторов. ^[105]

Искусственные нейронные сети

Искусственная нейронная сеть основана на наборе узлов, также известных как искусственные нейроны, которые в общих чертах моделируют нейроны биологического мозга. Она обучена распознавать закономерности; после обучения она может распознавать эти закономерности в новых данных. У неё есть вход, как минимум один скрытый слой узлов и выход. Каждый узел применяет функцию, и как только вес превышает заданный порог, данные передаются на следующий слой. Сеть обычно называется глубокой нейронной сетью, если она имеет как минимум два скрытых слоя. ^[105]

Алгоритмы обучения нейронных сетей используют локальный поиск для выбора весовых коэффициентов, которые обеспечат правильный результат для каждого входного значения во время обучения. Наиболее распространённым методом обучения является алгоритм обратного распространения ошибки. ^[106] Нейронные сети обучаются моделировать сложные взаимосвязи между входными и выходными данными и находить закономерности в данных. Теоретически нейронная сеть может обучиться любой функции. ^[107]

В нейронных сетях прямого распространения сигнал передается только в одном направлении. ^[108] Термин персептрон обычно относится к однослойной нейронной сети. ^[109] Напротив, глубокое обучение использует много слоев. ^[110] Рекуррентные нейронные сети (RNN) подают выходной сигнал обратно на вход, что позволяет кратковременно помнить предыдущие входные события. Сети с долговременной кратковременной памятью (LSTM) являются рекуррентными нейронными сетями, которые лучше сохраняют долговременные зависимости и менее чувствительны к проблеме исчезающего градиента. ^[111] Свёрточные нейронные сети (CNN) используют слои ядер для более эффективной обработки локальных шаблонов. Эта локальная обработка особенно важна при обработке изображений, где ранние слои CNN обычно идентифицируют простые локальные шаблоны, такие как края и кривые, а последующие слои обнаруживают более сложные шаблоны, такие как текстуры, и в конечном итоге целые объекты. ^[112]

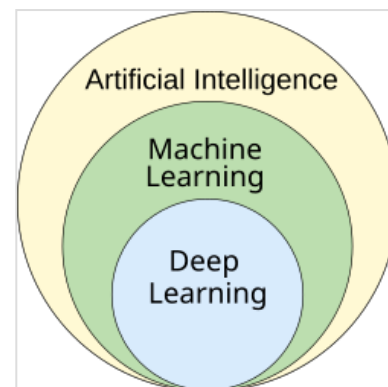


Нейронная сеть — это взаимосвязанная группа узлов, подобная обширной сети нейронов в человеческом мозге.

Глубокое обучение

Глубокое обучение использует несколько слоёв нейронов между входами и выходами сети. ^[110] Многочисленные слои позволяют постепенно извлекать более высокоуровневые признаки из необработанных входных данных. Например, при обработке изображений нижние слои могут определять контуры, в то время как верхние слои могут определять понятия, релевантные для человека, такие как цифры, буквы или лица. ^[114]

Глубокое обучение значительно повысило производительность программ во многих важных областях искусственного интеллекта, включая компьютерное зрение , распознавание речи , обработку естественного языка , классификацию изображений ^[115] и другие. Причина, по которой глубокое обучение так хорошо работает во многих приложениях, по состоянию на 2021 год неизвестна. ^[116] Внезапный успех глубокого обучения в 2012–2015 годах произошел не из-за какого- то нового открытия или теоретического прорыва (глубокие нейронные сети и обратное распространение были описаны многими людьми еще в 1950-х годах) ^[i] , а из-за двух факторов: невероятного увеличения мощности компьютеров (включая стократное увеличение скорости за счет перехода на графические процессоры) и доступности огромных объемов обучающих данных, особенно гигантских курируемых наборов данных, используемых для тестирования производительности, таких как ImageNet . ^[j]



Глубокое обучение – это подмножество машинного обучения , которое, в свою очередь, является подмножеством искусственного интеллекта. ^[113]

ГПТ

Генеративные предобученные преобразователи (GPT) – это большие языковые модели (LLM), которые генерируют текст на основе семантических связей между словами в предложениях. Текстовые модели GPT предварительно обучаются на большом корпусе текстов , которые могут быть взяты из Интернета. Предобучение заключается в прогнозировании следующего токена (токеном обычно является слово, часть слова или знак препинания). В ходе этого предобучения модели GPT накапливают знания об окружающем мире и затем могут генерировать текст, подобный человеческому, многократно предсказывая следующий токен. Как правило, последующая фаза обучения делает модель более правдивой, полезной и безвредной, обычно с помощью техники, называемой обучением с подкреплением на основе обратной связи с человеком (RLHF). Современные модели GPT склонны генерировать ложные сообщения, называемые « галлюцинациями ». Их можно уменьшить с помощью RLHF и качественных данных, но эта проблема усугубляется для систем рассуждений. ^[124] Такие системы используются в чат-ботах , которые позволяют людям задавать вопросы или запрашивать задание простым текстом. ^{[125] [126]}

Текущие модели и сервисы включают ChatGPT , Claude , Gemini , Copilot и Meta AI . ^[127] Мультимодальные модели GPT могут обрабатывать различные типы данных (модальности), такие как изображения, видео, звук и текст. ^[128]

Аппаратное и программное обеспечение

В конце 2010-х годов графические процессоры (GPU), которые все чаще разрабатывались с усовершенствованиями, специфическими для ИИ, и использовались со специализированным программным обеспечением TensorFlow , заменили ранее используемые центральные процессоры (CPU) в качестве основного средства для обучения крупномасштабных (коммерческих и академических) моделей машинного обучения . ^[129] Специализированные языки программирования , такие как Prolog , использовались в ранних исследованиях ИИ, ^[130] но языки программирования общего назначения , такие как Python, стали преобладающими. ^[131]

Плотность транзисторов в интегральных схемах, как было замечено, удваивается примерно каждые 18 месяцев — тенденция, известная как закон Мура , названный в честь соучредителя Intel Гордона Мура , который впервые его обнаружил. Усовершенствования графических процессоров происходили ещё быстрее ^[132] , эта тенденция иногда называется законом Хуанга ^[133] , названным в честь соучредителя и генерального директора Nvidia Дженсена Хуанга .

Приложения

Технология искусственного интеллекта и машинного обучения используется в большинстве основных приложений 2020-х годов, включая: поисковые системы (например, Google Search), таргетинг онлайн-рекламы , рекомендательные системы (предлагаемые Netflix , YouTube или Amazon), управление интернет-трафиком , целевую рекламу (AdSense , Facebook), виртуальные помощники (например, Siri или Alexa), автономные транспортные средства (включая дроны , ADAS и беспилотные автомобили), автоматический перевод языка (Microsoft Translator , Google Translate), распознавание лиц (FaceID от Apple или DeepFace от Microsoft и FaceNet от Google) и маркировку изображений (используется Facebook , Apple Photos и TikTok). Развертывание ИИ может контролироваться директором по автоматизации (CAO).

Здоровье и медицина

Применение ИИ в медицине и медицинских исследованиях может повысить качество ухода за пациентами и жизни. ^[134] В свете клятвы Гиппократа медицинские работники этически обязаны использовать ИИ, если приложения могут более точно диагностировать и лечить пациентов. ^[135]^[136]

Для медицинских исследований ИИ является важным инструментом для обработки и интеграции больших данных . Это особенно важно для разработки органовидов и тканевой инженерии , где микроскопическая визуализация является ключевым методом изготовления. ^[137] Было высказано предположение, что ИИ может преодолеть расхождения в финансировании, выделяемом на различные области исследований. ^[137]^[138] Новые инструменты ИИ могут углубить понимание биомедицинских значимых путей. Например, AlphaFold 2 (2021) продемонстрировал способность аппроксимировать за часы, а не за месяцы, трехмерную структуру белка . ^[139] В 2023 году сообщалось, что открытие лекарств под руководством ИИ помогло найти класс антибиотиков, способных убивать два разных типа устойчивых к лекарствам бактерий. ^[140] В 2024 году исследователи использовали машинное обучение для ускорения поиска лекарственных

средств для лечения болезни Паркинсона . Их целью было выявление соединений, блокирующих агрегацию альфа-синуклеина (белка, характерного для болезни Паркинсона). Им удалось ускорить процесс первичного скрининга в десять раз и снизить его стоимость в тысячу раз. ^[141]^[142]

Игры

Программы для игр использовались с 1950-х годов для демонстрации и тестирования самых передовых методов искусственного интеллекта. ^[143] Deep Blue стала первой компьютерной системой для игры в шахматы, которая победила действующего чемпиона мира по шахматам Гарри Каспарова 11 мая 1997 года. ^[144] В 2011 году в показательном матче викторины Jeopardy! система ответов на вопросы IBM Watson победила двух величайших чемпионов Jeopardy ! Брэда Раттера и Кена Дженинга со значительным отрывом. ^[145] В марте 2016 года AlphaGo выиграла 4 из 5 партий в го в матче с чемпионом го Ли Седолем , став первой компьютерной системой для игры в го, которая победила профессионального игрока в го без гандикапов . Затем, в 2017 году, она победила Кэ Цзе , который был лучшим игроком в го в мире. ^[146] Другие программы обрабатывают игры с неполной информацией , например, программа для игры в покер Pluribus . ^[147] DeepMind разработала все более общие модели обучения с подкреплением , такие как MuZero , которые можно обучить играть в шахматы, го или игры Atari . ^[148] В 2019 году AlphaStar от DeepMind достиг уровня гроссмейстера в StarCraft II , особенно сложной стратегической игре в реальном времени, которая требует неполного знания того, что происходит на карте. ^[149] В 2021 году агент ИИ участвовал в соревновании PlayStation Gran Turismo , победив четырех лучших в мире водителей Gran Turismo, используя глубокое обучение с подкреплением. ^[150] В 2024 году Google DeepMind представила SIMA, тип ИИ, способный автономно играть в девять ранее невиданных видеоигр в открытом мире , наблюдая за выводом на экран, а также выполняя короткие конкретные задачи в ответ на инструкции естественного языка. ^[151]

Математика

Большие языковые модели, такие как GPT-4 , Gemini , Claude , Llama или Mistral , все чаще используются в математике. Эти вероятностные модели универсальны, но также могут давать неправильные ответы в виде галлюцинаций . Иногда им нужна большая база данных математических задач для обучения, а также такие методы, как контролируемая тонкая настройка ^[152] или обученные классификаторы с аннотированными человеком данными для улучшения ответов на новые задачи и обучения на исправлениях. ^[153] Исследование, проведенное в феврале 2024 года, показало, что производительность некоторых языковых моделей для возможностей рассуждения при решении математических задач, не включенных в их тренировочные данные, была низкой, даже для задач с незначительными отклонениями от обученных данных. ^[154] Один из методов улучшения их производительности заключается в обучении моделей производить правильные шаги рассуждения , а не только правильный результат. ^[155] Alibaba Group разработала версию своих моделей Qwen под названием Qwen2-Math , которая достигла высочайших результатов в нескольких математических тестах, включая точность 84% на наборе данных MATH для математических задач соревнований. ^[156] В январе 2025 года Microsoft предложила технологию rStar-Math , которая использует поиск по дереву Монте-Карло и пошаговые рассуждения, что позволяет относительно небольшой языковой модели, такой как Qwen-7B, решить 53% задач AIME 2024 и 90% задач теста MATH. ^[157]

В качестве альтернативы были разработаны специализированные модели для решения математических задач с более высокой точностью результата, включая доказательство теорем, такие как *AlphaTensor* , *AlphaGeometry* , *AlphaProof* и *AlphaEvolve* ^[158] все от Google DeepMind , ^[159] *Llemma* от EleutherAI ^[160] или *Julius* ^[161] .

При использовании естественного языка для описания математических задач конвертеры могут преобразовывать такие подсказки в формальный язык, такой как Lean , для определения математических задач. Экспериментальная модель *Gemini Deep Think* напрямую воспринимает подсказки на естественном языке и завоевала золотую медаль на Международной математической олимпиаде 2025 года. ^[162]

Некоторые модели были разработаны для решения сложных задач и достижения хороших результатов в контрольных тестах, другие – для использования в качестве образовательных инструментов по математике. ^[163]

Топологическое глубокое обучение объединяет различные топологические подходы.

Финансы

Финансы являются одним из наиболее быстрорастущих секторов, где внедряются прикладные инструменты искусственного интеллекта: от розничного онлайн-банкинга до инвестиционного консультирования и страхования, где автоматизированные «роботы-консультанты» используются уже несколько лет. ^[164]

По мнению Николаса Фирцли, директора Всемирного форума по пенсиям и инвестициям , пока ещё слишком рано ожидать появления инновационных финансовых продуктов и услуг на основе искусственного интеллекта. Он утверждает, что «внедрение инструментов ИИ лишь ещё больше автоматизирует процессы, уничтожая десятки тысяч рабочих мест в банковской сфере, финансовом планировании и пенсионном консультировании, но я не уверен, что это приведёт к новой волне [например, сложных] пенсионных инноваций». ^[165]

Военный

Различные страны используют военные приложения ИИ. ^[166] Основные приложения улучшают командование и управление , связь, датчики, интеграцию и взаимодействие. ^[167] Исследования направлены на сбор и анализ разведданных, логистику, кибероперации, информационные операции, а также полуавтономные и автономные транспортные средства . ^[166] Технологии ИИ позволяют координировать датчики и исполнительные механизмы, обнаруживать и идентифицировать угрозы, отмечать позиции противника, захватывать цели , координировать и разрешать конфликты распределенных совместных огневых действий между сетевыми боевыми машинами, как управляемыми человеком, так и автономными . ^[167]

ИИ использовался в военных операциях в Ираке, Сирии, Израиле и Украине. ^[166] ^[168] ^[169] ^[170]

Генеративный ИИ

Генеративный искусственный интеллект (Generative AI, GenAI, ^[171] или GAI) — это подраздел искусственного интеллекта, который использует генеративные модели для создания текста, изображений, видео или других форм данных. ^[172] ^[173] ^[174] Эти модели изучают базовые шаблоны и структуры своих обучающих данных и используют их для создания новых данных ^[175] ^[176] на основе входных данных, которые часто поступают в виде подсказок на естественном языке. ^[177] ^[178]

Инструменты генеративного ИИ стали более распространенными после бума ИИ в 2020-х годах. Этот бум стал возможен благодаря улучшениям в глубоких нейронных сетях на основе трансформаторов, в частности, больших языковых моделях (LLM). Основные инструменты включают чат-ботов, таких как ChatGPT, Copilot, Gemini, Claude, Grok и DeepSeek; модели преобразования текста в изображения, такие как Stable Diffusion, Midjourney и DALL-E; и модели преобразования текста в видео, такие как Veo и Sora. ^[179] ^[180] ^[181] ^[182] ^[183] Технологические компании, разрабатывающие генеративный ИИ, включают OpenAI, xAI, Anthropic, Meta AI, Microsoft, Google, DeepSeek и Baidu. ^[177] ^[184] ^[185]



Винсент Ван Гог в акварели, созданной с помощью генеративного программного обеспечения искусственного интеллекта

Генеративный ИИ поднял множество этических вопросов и проблем управления, поскольку он может быть использован для киберпреступности или для обмана или манипулирования людьми с помощью фейковых новостей или дипфейков. ^[186] ^[187] Даже если он используется этично, он может привести к массовой замене человеческих рабочих мест. ^[188] Сами инструменты подвергались критике за нарушение законов об интеллектуальной собственности, поскольку они обучаются на работах, защищенных авторским правом. ^[189] Генеративный ИИ используется во многих отраслях, включая разработку программного обеспечения, ^[190] здравоохранение, ^[191] финансы, ^[192] развлечения, ^[193] обслуживание клиентов, ^[194] продажи и маркетинг, ^[195] искусство, письмо, ^[196] мода, ^[197] и дизайн продукта. ^[198]

Агенты

Агенты ИИ — это программные сущности, предназначенные для восприятия окружающей среды, принятия решений и выполнения действий автономно для достижения определенных целей. Эти агенты могут взаимодействовать с пользователями, своей средой или другими агентами. Агенты ИИ используются в различных приложениях, включая виртуальных помощников, чат-ботов, автономные транспортные средства, игровые системы и промышленную робототехнику. Агенты ИИ действуют в рамках ограничений своего программирования, доступных вычислительных ресурсов и аппаратных ограничений. Это означает, что они ограничены выполнением задач в пределах определенной области и имеют ограниченные возможности памяти и обработки. В реальных приложениях агенты ИИ часто сталкиваются с ограничениями по времени для принятия решений и выполнения действий. Многие агенты ИИ включают в себя алгоритмы

обучения, что позволяет им со временем улучшать свою производительность за счет опыта или обучения. Используя машинное обучение, агенты ИИ могут адаптироваться к новым ситуациям и оптимизировать свое поведение для поставленных задач. ^{[199] [200] [201]}

Сексуальность

Приложения ИИ в этой области включают в себя отслеживатели менструации и фертильности на основе ИИ, которые анализируют пользовательские данные для предоставления прогнозов, ^[202] секс-игрушки, интегрированные с ИИ (например, теледилдоника), ^[203] контент сексуального образования, созданный ИИ, ^[204] и агенты ИИ, которые имитируют сексуальных и романтических партнеров (например, *Replika*). ^[205] ИИ также используется для производства неконсенсуальной глубокой фэйковой порнографии , что вызывает серьезные этические и правовые проблемы. ^[206]

Технологии искусственного интеллекта также использовались для выявления случаев гендерного насилия в Интернете и сексуального насилия в отношении несовершеннолетних в Интернете. ^{[207] [208]}

Другие отраслевые задачи

Существуют также тысячи успешных приложений ИИ, используемых для решения конкретных задач в конкретных отраслях или учреждениях. Согласно опросу 2017 года, каждая пятая компания сообщила о внедрении ИИ в некоторые продукты или процессы. ^[209] В качестве примеров можно привести хранение энергии , медицинскую диагностику, военную логистику, приложения, прогнозирующие результаты судебных решений, внешнюю политику или управление цепочками поставок.

Применение ИИ для эвакуации и управления стихийными бедствиями растёт. ИИ используется для изучения закономерностей крупномасштабных и мелкомасштабных эвакуаций с использованием исторических данных GPS, видеозаписей и социальных сетей. Более того, ИИ может предоставлять информацию об условиях эвакуации в режиме реального времени. ^{[210] [211] [212]}

В сельском хозяйстве ИИ помогает фермерам повышать урожайность и выявлять области, требующие орошения, удобрений и обработки пестицидами. Агрономы используют ИИ для проведения исследований и разработок. ИИ используется для прогнозирования сроков созревания таких культур, как томаты, мониторинга влажности почвы, управления сельскохозяйственными роботами, проведения предиктивной аналитики , классификации эмоций скота и свиней, автоматизации теплиц, выявления болезней и вредителей, а также экономии воды.

Искусственный интеллект используется в астрономии для анализа всё большего объёма доступных данных и приложений, главным образом для «классификации, регрессии, кластеризации, прогнозирования, генерации, открытия и разработки новых научных идей». Например, он используется для открытия экзопланет, прогнозирования солнечной активности и различения сигналов и инструментальных эффектов в гравитационно-волновой астрономии. Кроме того, он может быть использован в космической деятельности, например, в космических исследованиях, включая анализ данных космических миссий, принятие научных решений в режиме реального времени для космических аппаратов, уклонение от космического мусора и обеспечение большей автономности.

Во время выборов в Индии в 2024 году было потрачено 50 миллионов долларов США на авторизованный контент, созданный с помощью искусственного интеллекта, в частности, на создание фейковых копий политиков-союзников (в том числе иногда и умерших) для лучшего взаимодействия с избирателями, а также на перевод речей на различные местные языки. ^[213]

Этика

ИИ имеет потенциальные преимущества и потенциальные риски. ^[216] ИИ может способствовать развитию науки и находить решения серьёзных проблем: Демис Хассабис из DeepMind надеется «разгадать проблему интеллекта, а затем использовать её для решения всего остального». ^[217] Однако по мере того, как использование ИИ стало повсеместным, было выявлено несколько непреднамеренных последствий и рисков. ^{[218] [219]} В производственных системах иногда не могут учитываться этические аспекты и предвзятость в процессах обучения ИИ, особенно когда алгоритмы ИИ по своей сути необъяснимы в глубоком обучении. ^[220]

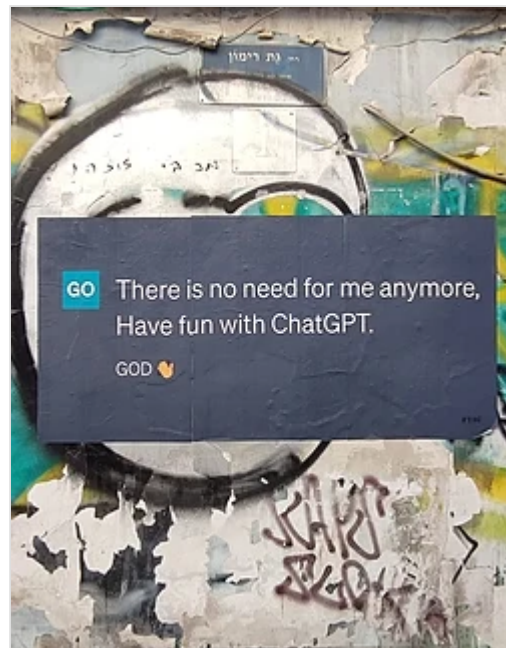
Риски и вред

Конфиденциальность и авторские права

Алгоритмы машинного обучения требуют больших объёмов данных. Методы, используемые для получения этих данных, вызывают опасения по поводу конфиденциальности , слежки и авторских прав .

Устройства и сервисы на базе искусственного интеллекта, такие как виртуальные помощники и продукты Интернета вещей, постоянно собирают персональные данные, что вызывает опасения по поводу навязчивого сбора данных и несанкционированного доступа со стороны третьих лиц. Утрата конфиденциальности усугубляется способностью искусственного интеллекта обрабатывать и объединять огромные объёмы данных, что потенциально приводит к созданию общества слежки, где действия отдельных лиц постоянно отслеживаются и анализируются без адекватных мер защиты и прозрачности.

Собранные конфиденциальные данные пользователей могут включать записи об онлайн-активности, данные геолокации, видео или аудио. ^[221] Например, для создания алгоритмов распознавания речи Amazon записала миллионы частных разговоров и позволила временным работникам прослушивать и расшифровывать некоторые из них. ^[222] Мнения об этой широко распространенной слежке варьируются от тех, кто считает ее необходимым злом, до тех, для кого это явно неэтично и является нарушением права на неприкосновенность частной жизни . ^[223]



Уличное искусство в Тель-Авиве

^{[214] [215]}

Разработчики ИИ утверждают, что это единственный способ предоставить ценные приложения, и разработали несколько методов, которые пытаются сохранить конфиденциальность, сохраняя при этом доступ к данным, такие как агрегация данных, деидентификация и дифференциальная конфиденциальность. ^[224] С 2016 года некоторые эксперты по конфиденциальности, такие как Синтия Дворк, начали рассматривать конфиденциальность с точки зрения справедливости. Брайан Кристиан написал, что эксперты перешли «от вопроса «что они знают» к вопросу «что они с этим делают»». ^[225]

Генеративный ИИ часто обучается на нелицензированных произведениях, защищенных авторским правом, в том числе в таких областях, как изображения или компьютерный код; затем полученные результаты используются в рамках обоснования «добросовестного использования». Эксперты расходятся во мнениях относительно того, насколько хорошо и при каких обстоятельствах это обоснование будет иметь силу в судах; значимыми факторами могут быть «цель и характер использования произведения, защищенного авторским правом» и «влияние на потенциальный рынок для этого произведения». ^{[226] [227]} Владельцы веб-сайтов, не желающие, чтобы их контент был удален, могут указать это в файле robots.txt. ^[228] В 2023 году ведущие авторы (включая Джона Гришэма и Джонатана Франзена) подали в суд на компании, разрабатывающие ИИ, за использование их работ для обучения генеративного ИИ. ^{[229] [230]} Другой обсуждаемый подход заключается в создании отдельной системы защиты творений, созданных ИИ, для обеспечения справедливой атрибуции и компенсации авторам-людям. ^[231]

Доминирование технологических гигантов

На рынке коммерческого ИИ доминируют такие крупные технологические компании, как Alphabet Inc., Amazon, Apple Inc., Meta Platforms и Microsoft. ^{[232] [233] [234]} Некоторые из этих игроков уже владеют подавляющим большинством существующей облачной инфраструктуры и вычислительной мощности центров обработки данных, что позволяет им еще больше укрепиться на рынке. ^{[235] [236]}

Потребности в электроэнергии и воздействие на окружающую среду

В январе 2024 года Международное энергетическое агентство (МЭА) опубликовало доклад «Электроэнергетика 2024: анализ и прогноз до 2026 года», в котором содержится прогноз потребления электроэнергии. ^[237] Это первый отчёт МЭА, содержащий прогнозы потребления электроэнергии центрами обработки данных и для искусственного интеллекта и криптовалют. В отчёте говорится, что спрос на электроэнергию для этих целей может удвоиться к 2026 году, при этом дополнительное потребление электроэнергии будет равно потреблению электроэнергии всей Японией. ^[238]

Огромное потребление энергии искусственным интеллектом (ИИ) обуславливает рост использования ископаемого топлива и может задержать закрытие устаревших угольных электростанций, генерирующих выбросы углерода. В США наблюдается стремительный рост строительства центров обработки данных, что превращает крупные технологические компании (например, Microsoft, Meta, Google, Amazon) в прожорливых потребителей электроэнергии. Прогнозируемое потребление электроэнергии настолько огромно, что существуют опасения, что оно будет удовлетворено независимо от источника. Поиск в ChatGPT предполагает использование в 10 раз большего количества электроэнергии, чем поиск в Google. Крупные компании спешат найти источники энергии – от ядерной и геотермальной энергии до термоядерного синтеза. Технологические компании утверждают,

что в долгосрочной перспективе ИИ станет более бережным к окружающей среде, но энергия им нужна уже сейчас. По данным технологических компаний, ИИ делает энергосистему более эффективной и «интеллектуальной», способствует развитию ядерной энергетики и отслеживает общие выбросы углерода. ^[239]

В исследовательском докладе *Goldman Sachs* 2024 года « *Центры обработки данных на базе искусственного интеллекта и грядущий всплеск спроса на электроэнергию в США* » говорится, что «спрос на электроэнергию в США, вероятно, будет расти, как никогда ранее...» и прогнозируется, что к 2030 году американские центры обработки данных будут потреблять 8% электроэнергии США по сравнению с 3% в 2022 году, что предвещает рост электроэнергетики различными способами. ^[240] Потребность центров обработки данных во всё большем количестве электроэнергии такова, что они могут перегрузить электросети. Крупные технологические компании возражают, что ИИ может быть использован для максимального использования электросети всеми. ^[241]

В 2024 году газета *Wall Street Journal* сообщила, что крупные компании, занимающиеся разработкой ИИ, начали переговоры с американскими поставщиками атомной энергии о поставках электроэнергии для центров обработки данных. В марте 2024 года Amazon приобрела центр обработки данных на атомной энергии в Пенсильвании за 650 миллионов долларов США. ^[242] Генеральный директор *Nvidia* Дженсен Хуанг заявил, что атомная энергетика — хороший вариант для центров обработки данных. ^[243]

В сентябре 2024 года *Microsoft* объявила о соглашении с *Constellation Energy* о повторном открытии атомной электростанции *Three Mile Island* , чтобы обеспечить *Microsoft* 100% всей электроэнергии, производимой станцией в течение 20 лет. Повторное открытие станции, которая пострадала от частичного расплавления ядерного реактора ее блока 2 в 1979 году, потребует от *Constellation* пройти строгие нормативные процедуры, которые будут включать обширную проверку безопасности со стороны *Комиссии по ядерному регулированию США* . В случае одобрения (это будет первый в истории США повторный ввод в эксплуатацию атомной станции) будет произведено более 835 мегаватт электроэнергии, что достаточно для 800 000 домов. Стоимость повторного открытия и модернизации оценивается в 1,6 миллиарда долларов США и зависит от налоговых льгот для ядерной энергетики, содержащихся в *Законое США о снижении инфляции 2022 года* . ^[244] Правительство США и штат Мичиган инвестируют почти 2 миллиарда долларов США в возобновление работы ядерного реактора *Palisades* на озере Мичиган. Закрытый с 2022 года, завод планируется открыть в октябре 2025 года. Центр чистой энергии *Crane Clean Energy Center* будет переименован в Центр чистой энергии *Crane* в честь Криса Крейна, сторонника ядерной энергетики и бывшего генерального директора *Exelon* , ответственного за отделение компании *Constellation* от *Exelon*. ^[245]

После последнего одобрения в сентябре 2023 года *Тайвань* приостановил одобрение центров обработки данных к северу от *Таоюаня* мощностью более 5 МВт в 2024 году из-за нехватки электроэнергии. ^[246] Тайвань намерен отказаться от ядерной энергетики к 2025 году. ^[246] С другой стороны, *Сингапур* ввел запрет на открытие центров обработки данных в 2019 году из-за электроэнергии, но в 2022 году снял этот запрет. ^[246]

Хотя большинство атомных электростанций в Японии были закрыты после аварии на АЭС «Фукусима» в 2011 году , согласно статье *Bloomberg* на японском языке за октябрь 2024 года, компания *Ubitus*, предоставляющая услуги облачных игр, в которой есть доля *Nvidia*, ищет землю в Японии рядом с атомной электростанцией для нового центра обработки

данных для генеративного ИИ. ^[247] Генеральный директор Ubitus Уэсли Куо заявил, что атомные электростанции являются наиболее эффективным, дешевым и стабильным источником энергии для ИИ. ^[247]

1 ноября 2024 года Федеральная комиссия по регулированию энергетики (FERC) отклонила заявку, поданную компанией Talen Energy на одобрение поставок электроэнергии с атомной электростанции Саскуэханна в центр обработки данных Amazon. ^[248] По словам председателя комиссии Вилли Л. Филлипса , это создает нагрузку на электросеть, а также вызывает значительную обеспокоенность по поводу перераспределения затрат для домохозяйств и других секторов бизнеса. ^[248]

В 2025 году, согласно докладу Международного энергетического агентства, выбросы парниковых газов , связанные с энергопотреблением ИИ, оценивались в 180 миллионов тонн. К 2035 году эти выбросы могут увеличиться до 300–500 миллионов тонн в зависимости от принятых мер. Это составляет менее 1,5% от общего объема выбросов энергетического сектора. Потенциал сокращения выбросов ИИ оценивается в 5% от общего объема выбросов энергетического сектора, но обратный эффект (например, если люди перейдут с общественного транспорта на беспилотные автомобили) может его снизить. ^[249]

Дезинформация

YouTube , Facebook и другие используют рекомендательные системы , чтобы направлять пользователей к большему количеству контента. Эти программы ИИ были поставлены целью максимизировать вовлеченность пользователей (то есть единственной целью было заставить людей смотреть). ИИ узнал, что пользователи склонны выбирать дезинформацию , теории заговора и крайне партийный контент, и, чтобы заставить их смотреть, ИИ рекомендовал больше этого. Пользователи также были склонны смотреть больше контента на одну и ту же тему, поэтому ИИ заводил людей в пузыри фильтров , где они получали несколько версий одной и той же дезинформации. ^[250] Это убедило многих пользователей в том, что дезинформация была правдой, и в конечном итоге подрывало доверие к институтам, СМИ и правительству. ^[251] Программа ИИ правильно научилась максимизировать свою цель, но результат был вреден для общества. После выборов в США в 2016 году крупные технологические компании предприняли некоторые шаги для смягчения проблемы. ^[252]

В начале 2020-х годов генеративный ИИ начал создавать изображения, аудио и тексты, которые практически неотличимы от настоящих фотографий, записей или человеческого текста ^[253] , а реалистичные видео, созданные ИИ, стали возможны в середине 2020-х годов. ^{[254] [255] [256]} Злоумышленники могут использовать эту технологию для создания огромных объемов дезинформации или пропаганды; ^[257] одним из таких потенциальных злонамеренных применений являются дипфейки для компьютерной пропаганды . ^[258] Пионер ИИ Джеффри Хинтон выразил обеспокоенность тем, что ИИ позволяет «авторитарным лидерам манипулировать своим электоратом» в больших масштабах, среди прочих рисков. ^[259]

Исследователи искусственного интеллекта из Microsoft , OpenAI , университетов и других организаций предложили использовать « учетные данные личности » как способ преодоления онлайн-обмана, обеспечиваемого моделями искусственного интеллекта. ^[260]

Алгоритмическая предвзятость и справедливость

Приложения машинного обучения будут предвзятыми ^[k], если они обучаются на предвзятых данных. ^[262] Разработчики могут не знать о существовании предвзятости. ^[263] Предвзятость может быть вызвана способом выбора данных для обучения и способом развертывания модели. ^[264] ^[262] Если предвзятый алгоритм используется для принятия решений, которые могут нанести серьезный вред людям (как это может быть в медицине, финансах, подборе персонала, жилищном строительстве или охране правопорядка), то этот алгоритм может стать причиной дискриминации. ^[265] Область справедливости изучает, как предотвратить вред от предвзятости алгоритмов.

28 июня 2015 года новая функция маркировки изображений в Google Фото ошибочно идентифицировала Джеки Алсина и его друга как «горилл», поскольку они были чернокожими. Система была обучена на наборе данных, содержащем очень мало изображений чернокожих людей ^[266], что является проблемой, известной как «несоответствие размера выборки» ^[267]. Google «исправил» эту проблему, запретив системе маркировать *что-либо* как «гориллу». Восемь лет спустя, в 2023 году, Google Фото по-прежнему не мог распознать гориллу, как и аналогичные продукты Apple, Facebook, Microsoft и Amazon. ^[268]

COMPAS — это коммерческая программа, широко используемая судами США для оценки вероятности рецидивизма обвиняемого. В 2016 году Джулия Энгвин из ProPublica обнаружила, что COMPAS демонстрирует расовую предвзятость, несмотря на то, что программе не сообщали расовую принадлежность обвиняемых. Хотя частота ошибок для белых и чернокожих была откалибрована одинаково и составляла ровно 61%, ошибки для каждой расы были разными — система постоянно переоценивала вероятность того, что чернокожий человек совершит повторное преступление, и недооценивала вероятность того, что белый человек не совершит повторное преступление. ^[269] В 2017 году несколько исследователей ^[1] показали, что для COMPAS математически невозможно учесть все возможные меры справедливости, когда базовые показатели повторных преступлений были разными для белых и чернокожих в данных. ^[271]

Программа может принимать предвзятые решения, даже если в данных явно не упоминается проблемный признак (например, «раса» или «пол»). Этот признак будет коррелировать с другими признаками (например, «адрес», «история покупок» или «имя»), и программа будет принимать на их основе те же решения, что и на основе «расы» или «пола». ^[272] Мориц Хардт сказал: «Самым убедительным фактом в этой области исследований является то, что справедливость через слепоту не работает». ^[273]

Критика COMPAS подчеркнула, что модели машинного обучения предназначены для составления «предсказаний», которые справедливы только в том случае, если мы предполагаем, что будущее будет похоже на прошлое. Если они обучены на данных, включающих результаты расистских решений в прошлом, модели машинного обучения должны предсказывать, что расистские решения будут приняты в будущем. Если приложение затем использует эти предсказания в качестве *рекомендаций*, некоторые из этих «рекомендаций», вероятно, будут расистскими. ^[274] Таким образом, машинное обучение не очень подходит для принятия решений в областях, где есть надежда, что будущее будет *лучше* прошлого. Оно носит описательный, а не предписывающий характер. ^[m]

Предвзятость и несправедливость могут остаться незамеченными, поскольку разработчики в подавляющем большинстве белые и мужчины: среди инженеров ИИ около 4% — чернокожие и 20% — женщины. ^[267]

Существуют различные противоречивые определения и математические модели справедливости. Эти понятия зависят от этических предположений и находятся под влиянием представлений об обществе. Одной из широких категорий является распределительная справедливость, которая фокусируется на результатах, часто идентифицируя группы и стремясь компенсировать статистические различия. Репрезентативная справедливость пытается гарантировать, что системы ИИ не усиливают негативные стереотипы или не делают определенные группы невидимыми. Процедурная справедливость фокусируется на процессе принятия решений, а не на результате. Наиболее релевантные понятия справедливости могут зависеть от контекста, в частности от типа приложения ИИ и заинтересованных сторон. Субъективность понятий предвзятости и справедливости затрудняет их операционализацию для компаний. Наличие доступа к чувствительным атрибутам, таким как раса или пол, также считается многими специалистами по этике ИИ необходимым для компенсации предвзятости, но это может противоречить антидискриминационным законам. ^[261]

На своей конференции 2022 года по справедливости, подотчетности и прозрачности (АСМ FAccT 2022) в Сеуле (Южная Корея) Ассоциация вычислительной техники представила и опубликовала выводы, в которых говорится, что до тех пор, пока не будет продемонстрировано отсутствие ошибок предвзятости в системах искусственного интеллекта и робототехники, они небезопасны, а использование самообучающихся нейронных сетей, обученных на обширных нерегулируемых источниках некорректных интернет-данных, следует ограничить. ^[276]

Отсутствие прозрачности

Многие системы искусственного интеллекта настолько сложны, что их разработчики не могут объяснить, как они принимают решения. ^[277] Особенно это касается глубоких нейронных сетей, в которых существует множество нелинейных связей между входными и выходными данными. Однако существуют некоторые популярные методы объяснимости. ^[278]

Невозможно быть уверенным в корректной работе программы, если никто не знает, как именно она работает. Было много случаев, когда программа машинного обучения проходила строгие тесты, но, тем не менее, обучалась чему-то иному, чем планировали программисты. Например, было обнаружено, что система, способная распознавать кожные заболевания лучше, чем медицинские работники, на самом деле имела выраженную тенденцию классифицировать изображения с линейкой как «раковые», поскольку изображения злокачественных новообразований обычно включают линейку для отображения масштаба. ^[279] Было обнаружено, что другая система машинного обучения, разработанная для эффективного распределения медицинских ресурсов, классифицировала пациентов с астмой как имеющих «низкий риск» смерти от пневмонии. Наличие астмы на самом деле является серьёзным фактором риска, но, поскольку пациенты с астмой обычно получали гораздо больше медицинской помощи, их смерть, согласно данным обучения, была относительно маловероятной. Корреляция между астмой и низким риском смерти от пневмонии была реальной, но вводящей в заблуждение. ^[280]

Люди, пострадавшие от решения алгоритма, имеют право на объяснение. ^[281] Например, врачи должны чётко и полно объяснять своим коллегам причины любого принятого ими решения. Ранние проекты Общего регламента Европейского союза по защите данных от 2016 года содержали прямое заявление о существовании такого права. ^[n] Эксперты отрасли отметили, что это нерешённая проблема, решения которой пока не видно. Регуляторы утверждали, что, тем не менее, вред реален: если проблема не имеет решения, инструменты не следует использовать. ^[282]

В 2014 году DARPA запустила программу XAI («Объяснимый искусственный интеллект»), чтобы попытаться решить эти проблемы. ^[283]

Несколько подходов направлены на решение проблемы прозрачности. SHAP позволяет визуализировать вклад каждого признака в выходные данные. ^[284] LIME может локально аппроксимировать выходные данные модели с помощью более простой, интерпретируемой модели. ^[285] Многозадачное обучение обеспечивает большое количество выходных данных в дополнение к целевой классификации. Эти другие выходные данные могут помочь разработчикам сделать вывод о том, чему научилась сеть. ^[286] Деконволюция , DeepDream и другие генеративные методы позволяют разработчикам увидеть, чему научились разные слои глубокой сети для компьютерного зрения, и создать выходные данные, которые могут подсказать, чему учится сеть. ^[287] Для генеративных предобученных преобразователей Anthropic разработала метод, основанный на обучении по словарю , который связывает шаблоны активации нейронов с понятными человеку понятиями. ^[288]

Злоумышленники и искусственный интеллект как оружие

Искусственный интеллект предоставляет ряд инструментов, которые полезны злоумышленникам , таким как авторитарные правительства , террористы , преступники или государства-изгои .

Смертоносное автономное оружие — это машина, которая обнаруживает, выбирает и поражает человеческие цели без человеческого контроля. ^[o] Широко доступные инструменты ИИ могут быть использованы злоумышленниками для разработки недорогого автономного оружия, и если они производятся в больших масштабах, они потенциально являются оружием массового поражения . ^[290] Даже при использовании в обычных боевых действиях они в настоящее время не могут надёжно выбирать цели и потенциально могут убить невинного человека . ^[290] В 2014 году 30 стран (включая Китай) поддержали запрет на автономное оружие в соответствии с Конвенцией Организации Объединённых Наций о конкретных видах обычного оружия , однако Соединённые Штаты и другие страны не согласились. ^[291] К 2015 году сообщалось, что более пятидесяти стран исследовали боевые роботы. ^[292]

Инструменты ИИ упрощают для авторитарных правительств эффективный контроль над своими гражданами несколькими способами. Распознавание лиц и голоса позволяет осуществлять широкомасштабное наблюдение . Машинное обучение , оперируя этими данными, может классифицировать потенциальных врагов государства и не давать им скрываться. Рекомендательные системы могут точно нацеливать пропаганду и дезинформацию для достижения максимального эффекта. Глубокие фейки и генеративный ИИ помогают в производстве дезинформации. Продвинутый ИИ может сделать авторитарное централизованное принятие решений более конкурентоспособным, чем либеральные и децентрализованные системы, такие как рынки . Это снижает стоимость и

сложность цифровой войны и передового шпионского ПО . ^[293] Все эти технологии доступны с 2020 года или ранее — системы распознавания лиц на основе ИИ уже используются для массового наблюдения в Китае. ^{[294] [295]}

Существует множество других способов, которыми ИИ может помочь злоумышленникам, некоторые из которых невозможно предвидеть. Например, машинное обучение ИИ позволяет создавать десятки тысяч токсичных молекул всего за несколько часов. ^[296]

Технологическая безработица

Экономисты часто подчеркивали риски увольнений из-за ИИ и рассуждали о безработице, если не будет адекватной социальной политики для полной занятости. ^[297]

В прошлом технологии имели тенденцию увеличивать, а не сокращать общую занятость, но экономисты признают, что с ИИ «мы находимся на неизведанной территории». ^[298] Опрос экономистов показал разногласия относительно того, вызовет ли растущее использование роботов и ИИ существенный рост долгосрочной безработицы , но в целом они согласны, что это может быть чистой выгодой, если рост производительности будет перераспределен . ^[299] Оценки риска различаются; например, в 2010-х годах Майкл Осборн и Карл Бенедикт Фрей оценили, что 47% рабочих мест в США находятся под «высоким риском» потенциальной автоматизации, в то время как в отчете ОЭСР только 9% рабочих мест в США были классифицированы как «высокорискованные». ^{[p] [301]} Методология спекуляций о будущих уровнях занятости подверглась критике за отсутствие доказательной базы и за то, что подразумевает, что технологии, а не социальная политика, создают безработицу, а не увольнения. ^[297] В апреле 2023 года сообщалось, что 70% рабочих мест для китайских иллюстраторов видеоигр были сокращены из-за генеративного искусственного интеллекта. ^{[302] [303]}

В отличие от предыдущих волн автоматизации, многие рабочие места среднего класса могут быть ликвидированы искусственным интеллектом. В 2015 году журнал *The Economist* заявил, что «опасения о том, что ИИ может сделать с работой «белых воротничков» то же, что паровая энергия сделала с работой «синих воротничков» во время промышленной революции, «стоит отнестись серьёзно». ^[304] К профессиям с высоким риском относятся профессии от помощников юристов до поваров фастфуда, при этом, вероятно, возрастет спрос на профессии, связанные с уходом, от личного медицинского обслуживания до священнослужителей. ^[305]

С первых дней развития искусственного интеллекта существовали аргументы, например, выдвинутые Джозефом Вайценбаумом , о том, должны ли задачи, которые могут быть выполнены компьютерами, на самом деле выполняться ими, учитывая разницу между компьютерами и людьми, а также между количественными расчетами и качественными, основанными на ценностях суждениями. ^[306]

Экзистенциальный риск

Утверждается, что ИИ станет настолько мощным, что человечество может безвозвратно потерять над ним контроль. Это может, как заметил физик Стивен Хокинг , « означать конец человеческой расы ». ^[307] Этот сценарий часто встречается в научной фантастике,

когда компьютер или робот внезапно обретает человеческое «самосознание» (или «чувство», или «сознание») и становится злобным персонажем. ^[9] Эти научно-фантастические сценарии вводят в заблуждение по нескольким причинам.

Во-первых, ИИ не обязательно должен обладать человеческим сознанием , чтобы представлять экзистенциальный риск. Современные программы ИИ получают конкретные цели и используют обучение и интеллект для их достижения. Философ Ник Бостром утверждал, что если дать *практически любую* цель достаточно мощному ИИ, он может решить уничтожить человечество для её достижения (он использовал пример с максимизатором скрепок). ^[309] Стюарт Рассел приводит пример домашнего робота, который пытается найти способ убить своего владельца, чтобы предотвратить его отключение от сети, рассуждая так: «Вы не сможете принести кофе, если вы мертвы». ^[310] Чтобы быть безопасным для человечества, сверхразум должен быть искренне разделяющим его мораль и ценности, чтобы он был «принципиально на нашей стороне». ^[311]

Во-вторых, Юваль Ной Харари утверждает, что ИИ не обязательно должен быть роботизированным телом или физическим контролем, чтобы представлять экзистенциальный риск. Важнейшие составляющие цивилизации не являются физическими. Такие вещи, как идеологии , право , правительство , деньги и экономика, построены на языке ; они существуют благодаря историям, в которые верят миллиарды людей. Нынешнее распространение дезинформации предполагает, что ИИ может использовать язык, чтобы убедить людей поверить во что угодно, даже совершить разрушительные действия. ^[312]

Мнения экспертов и представителей отрасли неоднозначны: значительная часть из них как обеспокоена, так и не обеспокоена риском, исходящим от возможного появления сверхразумного ИИ. ^[313] Такие личности, как Стивен Хокинг , Билл Гейтс и Илон Маск , ^[314] , а также пионеры в области ИИ, такие как Йошуа Бенджио , Стюарт Рассел , Демис Хассабис и Сэм Альтман , выражали обеспокоенность по поводу экзистенциального риска, исходящего от ИИ.

В мае 2023 года Джеффри Хинтон объявил о своей отставке из Google , чтобы иметь возможность «свободно говорить о рисках ИИ», не «думая о том, как это повлияет на Google». ^[315] Он, в частности, упомянул о рисках захвата ИИ ^[316] и подчеркнул, что для предотвращения худших последствий разработка правил безопасности потребует сотрудничества между теми, кто конкурирует в использовании ИИ. ^[317]

В 2023 году многие ведущие эксперты в области искусственного интеллекта одобрили совместное заявление о том, что «Снижение риска вымирания из-за искусственного интеллекта должно стать глобальным приоритетом наряду с другими рисками общественного масштаба, такими как пандемии и ядерная война». ^[318]

Некоторые другие исследователи были настроены более оптимистично. Пионер ИИ Юрген Шмидхубер не подписал совместное заявление, подчеркнув, что в 95% случаев исследования ИИ направлены на то, чтобы сделать «человеческую жизнь более продолжительной, здоровой и легкой». ^[319] Хотя инструменты, которые сейчас используются для улучшения жизни, могут использоваться и злоумышленниками, «они также могут быть использованы против злоумышленников». ^{[320] [321]} Эндрю Нг также утверждал, что «ошибочно поддаваться апокалипсису вокруг ИИ, и что регулирующие органы, которые это сделают, только выиграют от корыстных интересов». ^[322] Янн Лекун

«высмеивает антиутопические сценарии своих коллег о чрезмерной дезинформации и даже, в конечном итоге, о вымирании человечества». ^[323] В начале 2010-х годов эксперты утверждали, что риски слишком отдаленного будущего, чтобы оправдать исследования, или что люди будут представлять ценность с точки зрения сверхразумной машины. ^[324] Однако после 2016 года изучение текущих и будущих рисков и возможных решений стало серьезной областью исследований. ^[325]

Этические машины и выравнивание

Дружественный ИИ — это машины, которые изначально проектировались с целью минимизировать риски и принимать решения, выгодные людям. Элиезер Юджовски , автор этого термина, утверждает, что разработка дружелюбного ИИ должна стать более приоритетной задачей для исследований: это может потребовать значительных инвестиций и должно быть завершено до того, как ИИ станет экзистенциальным риском. ^[326]

Машины с интеллектом потенциально способны использовать свой интеллект для принятия этических решений. Область машинной этики предоставляет машинам этические принципы и процедуры для решения этических дилемм. ^[327] Область машинной этики также называется вычислительной моралью ^[327] и была основана на симпозиуме AAAI в 2005 году. ^[328]

Другие подходы включают «искусственных моральных агентов» Уэнделла Уоллаха ^[329] и три принципа Стюарта Дж. Рассела для разработки доказуемо полезных машин. ^[330]

Открытый исходный код

Активные организации в сообществе открытого исходного кода ИИ включают Hugging Face , ^[331] Google , ^[332] EleutherAI и Meta . ^[333] Различные модели ИИ, такие как Llama 2 , Mistral или Stable Diffusion , стали открытыми, ^[334]^[335] что означает, что их архитектура и обученные параметры («веса») общедоступны. Модели с открытым весом можно свободно настраивать , что позволяет компаниям специализировать их на собственных данных и для собственных вариантов использования. ^[336] Модели с открытым весом полезны для исследований и инноваций, но их также можно использовать не по назначению. Поскольку их можно настраивать точно, любая встроенная мера безопасности, такая как возражение против вредоносных запросов, может быть обучена до тех пор, пока она не станет неэффективной. Некоторые исследователи предупреждают, что будущие модели ИИ могут развить опасные возможности (например, потенциально значительно способствовать биотерроризму), и что после публикации в Интернете их невозможно будет удалить везде, где это необходимо. Они рекомендуют проводить предварительный аудит и анализ затрат и выгод. ^[337]

Фреймворки

Проекты искусственного интеллекта могут руководствоваться этическими соображениями на этапах проектирования, разработки и внедрения системы ИИ. Фреймворк ИИ, такой как «Фреймворк заботы и действия», разработанный Институтом Алана Тьюринга и основанный на ценностях SUM, выделяет четыре основных этических аспекта, которые определяются следующим образом: ^[338]^[339]

- **Уважайте** достоинство отдельных людей
- **Общайтесь** с другими людьми искренне, открыто и инклюзивно.
- **Забота** о благополучии каждого
- **Защищать** социальные ценности, справедливость и общественные интересы

Другие разработки в области этических рамок включают решения, принятые на конференции Асиломар , Монреальскую декларацию об ответственном ИИ и инициативу IEEE по этике автономных систем, среди прочих; ^[340] однако эти принципы не лишены критики, особенно в отношении людей, выбранных для участия в этих рамках. ^[341]

Для повышения благополучия людей и сообществ, на которые влияют эти технологии, необходимо учитывать социальные и этические последствия на всех этапах проектирования, разработки и внедрения систем ИИ, а также сотрудничество между специалистами по работе с данными, менеджерами по продуктам, инженерами по работе с данными, экспертами в предметной области и менеджерами по доставке. ^[342]

Британский институт безопасности искусственного интеллекта (ИИ) в 2024 году выпустил набор инструментов для тестирования под названием «Inspect» для оценки безопасности ИИ, доступный по лицензии MIT с открытым исходным кодом. Набор инструментов доступен бесплатно на GitHub и может быть улучшен с помощью сторонних пакетов. Он может использоваться для оценки моделей ИИ в различных областях, включая базовые знания, способность к рассуждению и автономные возможности. ^[343]

Регулирование

Регулирование искусственного интеллекта представляет собой разработку политик и законов государственного сектора для продвижения и регулирования ИИ; поэтому оно связано с более широким регулированием алгоритмов. ^[344] Ландшафт регулирования и политики для ИИ является новой проблемой в юрисдикциях по всему миру. ^[345] Согласно Индексу ИИ в Стэнфорде , ежегодное количество законов, связанных с ИИ, принятых в 127 странах, участвовавших в исследовании, подскочило с одного, принятого в 2016 году, до 37, принятых только в 2022 году. ^[346] ^[347] В период с 2016 по 2020 год более 30 стран приняли специальные стратегии для ИИ. ^[348]

Большинство государств-членов ЕС выпустили национальные стратегии в области ИИ, как и Канада, Китай, Индия, Япония, Маврикий, Российская Федерация, Саудовская Аравия, Объединенные Арабские Эмираты, США и Вьетнам. Другие находились в процессе разработки собственной стратегии в области ИИ, включая Бангладеш, Малайзию и Тунис. ^[348] Глобальное партнерство по искусственному интеллекту было запущено в июне 2020 года, заявив о необходимости разработки ИИ в соответствии с правами человека и демократическими ценностями, чтобы обеспечить общественное доверие к технологии. ^[348] Генри Киссинджер , Эрик Шмидт и Дэниел Хуттенлохер опубликовали совместное заявление в ноябре 2021 года, призвав к созданию правительственной комиссии для регулирования ИИ. ^[349] В 2023 году лидеры OpenAI



Первый всемирный саммит по безопасности ИИ прошел в Великобритании в ноябре 2023 года, на нем была принята декларация, призывающая к международному сотрудничеству.

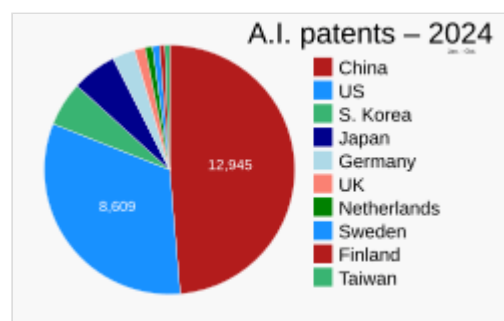
опубликовали рекомендации по управлению сверхинтеллектом, которое, по их мнению, может произойти менее чем через 10 лет. ^[350] В 2023 году Организация Объединенных Наций также создала консультативный орган для предоставления рекомендаций по управлению ИИ; в состав органа входят руководители технологических компаний, государственные служащие и ученые. ^[351] 1 августа 2024 года вступил в силу Закон ЕС об искусственном интеллекте , устанавливающий первое всеобъемлющее регулирование ИИ на уровне ЕС. ^[352] В 2024 году Совет Европы разработал первый международный юридически обязывающий договор об искусственном интеллекте, названный «Рамочная конвенция об искусственном интеллекте и правах человека, демократии и верховенстве права ». Он был принят Европейским союзом, Соединенными Штатами, Соединенным Королевством и другими странами. ^[353]

В опросе Ipsos 2022 года отношение к ИИ сильно различалось в зависимости от страны; 78% граждан Китая, но только 35% американцев согласились с тем, что «продукты и услуги, использующие ИИ, имеют больше преимуществ, чем недостатков». ^[346] Опрос Reuters /Ipsos 2023 года показал, что 61% американцев согласны, а 22% не согласны с тем, что ИИ представляет опасность для человечества. ^[354] В опросе Fox News 2023 года 35% американцев считали «очень важным», а ещё 41% считали «довольно важным» регулирование ИИ федеральным правительством, в то время как 13% ответили «не очень важно», а 8% — «совсем не важно». ^[355] ^[356]

В ноябре 2023 года в Блетчли-Парке (Великобритания) состоялся первый всемирный саммит по безопасности ИИ , на котором обсуждались краткосрочные и долгосрочные риски ИИ, а также возможность создания обязательных и добровольных нормативных рамок. ^[357] 28 стран, включая США, Китай и Европейский союз, опубликовали декларацию в начале саммита, призывающую к международному сотрудничеству для управления вызовами и рисками искусственного интеллекта. ^[358] ^[359] В мае 2024 года на саммите ИИ в Сеуле 16 глобальных компаний, занимающихся технологиями ИИ, согласились взять на себя обязательства по обеспечению безопасности при разработке ИИ. ^[360] ^[361]

История

Изучение механических или «формальных» рассуждений началось с философов и математиков древности. Изучение логики привело непосредственно к теории вычислений Алана Тьюринга , которая предполагала, что машина, перетасовывая такие простые символы, как «0» и «1», может имитировать любую мыслимую форму математических рассуждений. ^[363] ^[364] Это, наряду с одновременными открытиями в кибернетике , теории информации и нейробиологии , привело исследователей к рассмотрению возможности создания «электронного мозга». ^[r] Они разработали несколько направлений исследований, которые стали частью ИИ, ^[366] , такие как проект Маккалока и Питтса для «искусственных нейронов» в 1943 году, ^[117] и влиятельная статья Тьюринга 1950 года «



В 2024 году патенты на ИИ в Китае и США составили более трех четвертей от общего числа патентов на ИИ в мире. ^[362] Хотя у Китая было больше патентов на ИИ, в США было на 35% больше патентов на компанию, подавшую заявку на патент на ИИ, чем в Китае. ^[362]

Вычислительная техника и интеллект », которая представила тест Тьюринга и показала, что «машинный интеллект» правдоподобен. ^{[367] [364]}

Область исследований искусственного интеллекта была основана на семинаре в Дартмутском колледже в 1956 году. ^{[s] [6]} Участники семинара стали лидерами исследований искусственного интеллекта в 1960-х годах. ^[t] Они и их студенты создали программы, которые пресса описала как «поразительные»: ^[u] компьютеры изучали стратегии игры в шашки , решали текстовые задачи по алгебре, доказывали логические теоремы и говорили по-английски. ^{[v] [7]} Лаборатории искусственного интеллекта были созданы в ряде британских и американских университетов в конце 1950-х и начале 1960-х годов. ^[364]

Исследователи 1960-х и 1970-х годов были убеждены, что их методы в конечном итоге приведут к созданию машины с общим интеллектом , и считали это целью своей области. ^[371] В 1965 году Герберт Саймон предсказал, что «машины будут способны в течение двадцати лет выполнять любую работу, которую может выполнить человек». ^[372] В 1967 году Марвин Мински согласился, написав, что «в течение одного поколения... проблема создания «искусственного интеллекта» будет в значительной степени решена». ^[373] Однако они недооценили сложность проблемы. ^[w] В 1974 году правительства США и Великобритании прекратили поисковые исследования в ответ на критику сэра Джеймса Лайтхилла ^[375] и постоянное давление со стороны Конгресса США с требованием финансировать более продуктивные проекты . ^[376] Книга Мински и Паперта «*Перцептроны*» была понята как доказательство того, что искусственные нейронные сети никогда не будут полезны для решения реальных задач, тем самым полностью дискредитируя этот подход. ^[377] Затем последовала «зима ИИ » — период, когда было трудно получить финансирование для проектов ИИ. ^[9]

В начале 1980-х годов исследования в области ИИ были возрождены коммерческим успехом экспертных систем ^[378] , формы программы ИИ, которая имитировала знания и аналитические навыки экспертов-людей. К 1985 году рынок ИИ достиг более миллиарда долларов. В то же время японский проект пятого поколения компьютеров вдохновил правительства США и Великобритании возобновить финансирование академических исследований ^[8] . Однако, начиная с краха рынка машин Lisp в 1987 году, ИИ снова оказался в дурной славе, и началась вторая, более продолжительная зима. ^[10]

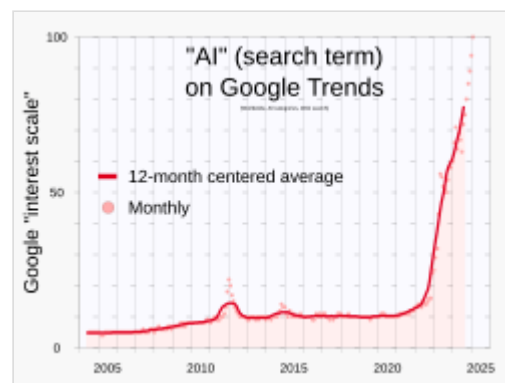
До этого момента большая часть финансирования ИИ направлялась на проекты, которые использовали символы высокого уровня для представления ментальных объектов , таких как планы, цели, убеждения и известные факты. В 1980-х годах некоторые исследователи начали сомневаться, что этот подход сможет имитировать все процессы человеческого познания, особенно восприятие , робототехнику , обучение и распознавание образов , ^[379] и начали изучать «субсимволические» подходы. ^[380] Родни Брукс отверг «представление» в целом и сосредоточился непосредственно на разработке машин, которые двигаются и выживают. ^[x] Джуда Перл , Лотфи Заде и другие разработали методы, которые обрабатывали неполную и неопределенную информацию, делая разумные догадки, а не точную логику. ^{[87] [385]} Но самым важным событием стало возрождение «коннекционизма » , включая исследования нейронных сетей, проведенные Джеффри Хинтоном и другими. ^[386] В 1990 году Ян Лекун успешно продемонстрировал, что сверточные нейронные сети могут распознавать рукописные цифры, что стало первым из многих успешных применений нейронных сетей. ^[387]

В конце 1990-х и начале 21-го века искусственный интеллект (ИИ) постепенно восстановил свою репутацию, применяя формальные математические методы и находя конкретные решения для конкретных задач. Эта «узкая» и «формальная» направленность позволила исследователям получать проверяемые результаты и сотрудничать с другими областями (такими как статистика, экономика и математика). [388] К 2000 году решения, разработанные исследователями ИИ, стали широко использоваться, хотя в 1990-х годах их редко называли «искусственным интеллектом» (тенденция, известная как эффект ИИ). [389] Однако некоторые ученые-исследователи стали беспокоиться, что ИИ больше не преследует свою первоначальную цель — создание универсальных, полностью интеллектуальных машин. Примерно в 2002 году они основали подраздел искусственного интеллекта общего назначения (или «ИИО»), в котором к 2010-м годам появилось несколько хорошо финансируемых институтов. [68]

Глубокое обучение стало доминировать в отраслевых бенчмарках в 2012 году и было принято во всей области. [11] Для многих конкретных задач другие методы были заброшены. [y] Успех глубокого обучения был основан как на усовершенствованиях оборудования (более быстрые компьютеры, [391] графические процессоры, облачные вычисления [392]), так и на доступе к большим объемам данных [393] (включая курируемые наборы данных, [392] такие как ImageNet). Успех глубокого обучения привел к огромному росту интереса и финансирования в области ИИ. [z] Объем исследований в области машинного обучения (измеряемый общим числом публикаций) увеличился на 50 % в период с 2015 по 2019 год. [348]

В 2016 году вопросы справедливости и злоупотребления технологиями оказались в центре внимания на конференциях по машинному обучению, число публикаций значительно увеличилось, финансирование стало доступным, и многие исследователи переориентировали свою карьеру на эти вопросы. Проблема согласования стала серьёзной областью академических исследований. [325]

В конце 2010-х и начале 2020-х годов компании, занимающиеся разработкой ИИ, начали выпускать программы, вызвавшие огромный интерес. В 2015 году AlphaGo, разработанная DeepMind, победила чемпиона мира по игре в го. Программа обучала только правилам игры и самостоятельно разрабатывала стратегию. GPT-3 — это большая языковая модель, выпущенная OpenAI в 2020 году и способная генерировать высококачественный текст, подобный человеческому. [394] ChatGPT, запущенный 30 ноября 2022 года, стал самым быстрорастущим потребительским программным приложением в истории, набрав более 100 миллионов пользователей за два месяца. [395] Он ознаменовал то, что широко считается годом прорыва ИИ, сделав его достоянием общественности. [396] Эти и другие программы вызвали агрессивный бум ИИ, когда крупные компании начали вкладывать миллиарды долларов в исследования ИИ. По данным AI Impacts, в 2022 году только в США в искусственный интеллект ежегодно инвестировалось около 50 миллиардов долларов США, а около 20% выпускников докторских программ в области компьютерных наук в США специализируются на искусственном интеллекте. [397] В 2022 году в США существовало около 800 000 вакансий,



Количество поисковых запросов Google по термину «ИИ» увеличилось в 2022 году.

связанных с искусственным интеллектом. ^[398] Согласно исследованию PitchBook, 22% новых стартапов, получивших финансирование в 2024 году, заявляли, что являются компаниями, работающими в сфере искусственного интеллекта. ^[399]

Философия

Философские дебаты исторически были направлены на определение природы интеллекта и того, как создавать интеллектуальные машины. ^[400] Другим важным вопросом было то, могут ли машины обладать сознанием, и связанные с этим этические последствия. ^[401] Многие другие темы философии имеют отношение к ИИ, такие как эпистемология и свободная воля . ^[402] Стремительный прогресс усилил общественные дискуссии о философии и этике ИИ . ^[401]

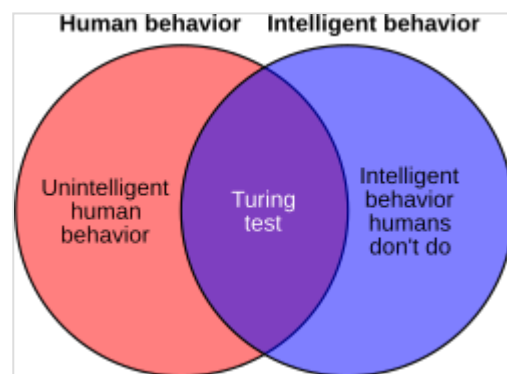
Определение искусственного интеллекта

Алан Тьюринг писал в 1950 году: «Я предлагаю рассмотреть вопрос „могут ли машины мыслить?“» ^[403]. Он посоветовал изменить вопрос с вопроса о том, «думает ли машина», на вопрос о том, «может ли машина проявлять разумное поведение». ^[403] . Он разработал тест Тьюринга , который измеряет способность машины имитировать человеческий разговор. ^[367]. Поскольку мы можем только наблюдать за поведением машины, неважно, думает ли она «на самом деле» или у неё буквально есть «разум». Тьюринг отмечает, что мы не можем определить эти вещи относительно других людей, но «обычно существует вежливое соглашение, что все думают». ^[404].

Рассел и Норвиг согласны с Тьюрингом в том, что интеллект следует определять с точки зрения внешнего поведения, а не внутренней структуры. ^[1] Однако они критически считают, что тест требует, чтобы машина имитировала людей. «Тексты по авиационной технике», — писали они, — не определяют цель этой области как создание „машин, которые летают настолько точно, как голуби“, что могут обманывать других голубей “ » . ^[406] Основатель искусственного интеллекта Джон Маккарти согласился, написав, что «Искусственный интеллект, по определению, не является имитацией человеческого интеллекта». ^[407]

Маккарти определяет интеллект как «вычислительную часть способности достигать целей в мире». ^[408]

Другой основатель ИИ, Марвин Мински , аналогичным образом описывает его как «способность решать сложные проблемы». ^[409] Ведущий учебник по ИИ определяет его как изучение агентов, которые воспринимают окружающую среду и предпринимают действия, которые максимизируют их шансы на достижение поставленных целей. ^[1] Эти определения рассматривают интеллект с точки зрения четко определенных проблем с четко определенными решениями, где как сложность проблемы, так и производительность программы являются прямыми мерами «интеллекта» машины — и никакое другое философское обсуждение не требуется или даже может быть невозможным.



Тест Тьюринга может предоставить некоторые доказательства наличия интеллекта, но он наказывает за нечеловеческое разумное поведение. ^[405]

Другое определение было принято компанией Google ^[410], крупным специалистом в области искусственного интеллекта. Это определение рассматривает способность систем синтезировать информацию как проявление интеллекта, аналогично тому, как это определяется в биологическом интеллекте.

В результате множества определений учёные начали критически анализировать и упорядочивать сам дискурс об ИИ ^[411], включая обсуждение многочисленных нарративов и мифов об ИИ, встречающихся в общественных, политических и академических дискуссиях. ^[412] Аналогичным образом, на практике некоторые авторы предполагают, что термин «ИИ» часто используется слишком широко и расплывчато. Это поднимает вопрос о том, где следует провести границу между ИИ и классическими алгоритмами, ^[413] поскольку многие компании в начале 2020-х годов, во время бума ИИ, использовали этот термин как модное маркетинговое словечко, часто даже если они «фактически не использовали ИИ материально». ^[414]

Ведутся споры о том, демонстрируют ли большие языковые модели настоящий интеллект или просто имитируют его, имитируя человеческий текст. ^[415]

Оценка подходов к ИИ

На протяжении большей части истории исследований в области ИИ не существовало устоявшейся объединяющей теории или парадигмы. ^[aa] Беспрецедентный успех статистического машинного обучения в 2010-х годах затмил все другие подходы (настолько, что некоторые источники, особенно в деловом мире, используют термин «искусственный интеллект» для обозначения «машинного обучения с использованием нейронных сетей»). Этот подход в основном субсимволический, «мягкий» и узкий. Критики утверждают, что будущим поколениям исследователей ИИ, возможно, придётся пересмотреть эти вопросы.

Символический ИИ и его ограничения

Символический ИИ (или «GOFAI») ^[417] имитировал высокоуровневое сознательное мышление, используемое людьми при решении головоломок, формулировании юридических рассуждений и выполнении математических задач. Он добился больших успехов в решении «интеллектуальных» задач, таких как алгебра или тесты на IQ. В 1960-х годах Ньюэлл и Саймон выдвинули гипотезу физических символьных систем: «Физическая символьная система обладает необходимыми и достаточными средствами для выполнения общих интеллектуальных действий». ^[418]

Однако символический подход оказался неэффективным при решении многих задач, которые люди легко решают, таких как обучение, распознавание объекта или рассуждения, основанные на здравом смысле. Парадокс Моравца заключается в открытии того, что высокоуровневые «интеллектуальные» задачи легко решаются ИИ, тогда как низкоуровневые «инстинктивные» задачи оказались чрезвычайно сложными. ^[419] Философ Губерт Дрейфус с 1960-х годов утверждал, что человеческие знания основаны на бессознательных инстинктах, а не на сознательном манипулировании символами, и на «чувстве» ситуации, а не на явном символическом знании. ^[420] Хотя его аргументы были высмеяны и проигнорированы, когда они были впервые представлены, в конечном итоге исследователи ИИ согласились с ним. ^{[ab] [16]}

Проблема не решена: субсимволическое мышление может совершать многие из тех же непостижимых ошибок, что и человеческая интуиция, например, алгоритмическую предвзятость . Критики, такие как Ноам Хомский, утверждают, что для достижения общего уровня интеллекта всё ещё необходимо продолжать исследования символического ИИ [422] [423], отчасти потому, что субсимволический ИИ — это шаг в сторону от объяснимого ИИ : может быть сложно или невозможно понять, почему современная статистическая программа ИИ приняла то или иное решение. Развивающееся направление нейросимволического искусственного интеллекта пытается объединить эти два подхода.

Аккуратный против неряшливого

«Чистоплотные» надеются, что интеллектуальное поведение описывается с помощью простых и элегантных принципов (таких как логика , оптимизация или нейронные сети). «Неряхи» ожидают, что это обязательно потребует решения большого количества несвязанных между собой задач. Чистоплотные защищают свои программы, опираясь на теоретическую строгость, неряхи же полагаются в основном на инкрементальное тестирование, чтобы проверить их работоспособность. Этот вопрос активно обсуждался в 1970-х и 1980-х годах [424] , но в конечном итоге был признан неактуальным. Современный ИИ обладает элементами обоих.

Мягкие и жесткие вычисления

Нахождение доказуемо правильного или оптимального решения является неразрешимой задачей для многих важных задач. [15] Мягкие вычисления представляют собой набор методов, включая генетические алгоритмы , нечёткую логику и нейронные сети, которые допускают неточность, неопределённость, частичную истинность и аппроксимацию. Мягкие вычисления появились в конце 1980-х годов, и большинство успешных программ искусственного интеллекта в XXI веке являются примерами мягких вычислений с нейронными сетями.

Узкий и общий ИИ

Исследователи ИИ разделились во мнениях относительно того, следует ли им напрямую стремиться к достижению целей общего искусственного интеллекта и сверхинтеллекта или решать как можно больше конкретных задач (узкий ИИ) в надежде, что эти решения косвенно приведут к достижению долгосрочных целей области. [425] [426] Общий интеллект сложно определить и измерить, и современный ИИ добился более проверяемых успехов, сосредоточившись на конкретных проблемах с конкретными решениями. Подраздел общего искусственного интеллекта изучает исключительно эту область.

Машинное сознание, ощущения и разум

В философии разума нет устоявшегося консенсуса относительно того, может ли машина обладать разумом , сознанием и ментальными состояниями в том же смысле, что и человек. Этот вопрос касается внутреннего опыта машины, а не её внешнего поведения. В традиционных исследованиях в области искусственного интеллекта этот вопрос неактуален, поскольку он не влияет на цели данной области: создание машин, способных решать задачи с помощью интеллекта. Рассел и Норвиг добавляют, что «дополнительный проект по созданию машины, обладающей сознанием, подобным человеческому, — это не тот, к

которому мы готовы». ^[427] Однако этот вопрос стал центральным в философии разума. Он также, как правило, является центральным вопросом, обсуждаемым в художественной литературе об искусственном интеллекте .

Сознание

Дэвид Чалмерс выделил две проблемы в понимании разума, которые он назвал «трудной» и «лёгкой» проблемами сознания. ^[428] Легкая проблема заключается в понимании того, как мозг обрабатывает сигналы, строит планы и контролирует поведение. Сложная проблема заключается в объяснении того, как это *ощущается* или почему это вообще должно ощущаться как что-то, предполагая, что мы правы, полагая, что это действительно ощущается как что-то (иллюзионизм сознания Деннета утверждает, что это иллюзия). В то время как человеческую обработку информации легко объяснить, субъективный человеческий опыт объяснить сложно. Например, легко представить себе дальтоника, который научился определять, какие объекты в его поле зрения красные, но неясно, что требуется, чтобы этот человек *знал, как выглядит красный цвет* . ^[429]

Вычислительность и функционализм

Вычислительный подход (Computationalism) — это позиция в философии сознания , согласно которой человеческий разум представляет собой систему обработки информации, а мышление — форму вычислений. Вычислительный подход утверждает, что отношения между разумом и телом аналогичны или идентичны отношениям между программным обеспечением и оборудованием и, таким образом, могут быть решением проблемы «разум—тело» . Эта философская позиция была вдохновлена работами исследователей искусственного интеллекта и когнитивных учёных в 1960-х годах и первоначально была предложена философами Джерри Фодором и Хилари Патнэмом . ^[430]

Философ Джон Сёрл охарактеризовал эту позицию как «сильный ИИ»: «Соответствующим образом запрограммированный компьютер с правильными входами и выходами будет обладать разумом в том же смысле, в каком им обладают люди». ^[ac] Сёрл оспаривает это утверждение с помощью своего аргумента китайской комнаты , который пытается показать, что даже компьютер, способный идеально имитировать человеческое поведение, не будет обладать разумом. ^[434]

Благосостояние и права ИИ

Трудно или невозможно достоверно оценить, является ли развитый ИИ разумным (обладает способностью чувствовать), и если да, то в какой степени. ^[435] Но если существует значительная вероятность того, что данная машина может чувствовать и страдать, то она может иметь право на определенные права или меры защиты благосостояния, аналогично животным. ^[436]^[437] Разумность (набор способностей, связанных с высоким интеллектом, таких как проницательность или самосознание) может предоставить еще одну моральную основу для прав ИИ. ^[436] Права роботов также иногда предлагаются как практический способ интеграции автономных агентов в общество. ^[438]

В 2017 году Европейский союз рассматривал возможность предоставления «электронной личности» некоторым наиболее эффективным системам искусственного интеллекта. Подобно правовому статусу компаний, это наделило бы их правами, но и ответственностью. ^[439] В 2018 году критики утверждали, что предоставление прав системам искусственного

интеллекта принижало бы важность прав человека и что законодательство должно быть ориентировано на потребности пользователей, а не на спекулятивные футуристические сценарии. Они также отмечали, что роботам не хватает автономности, чтобы самостоятельно участвовать в жизни общества. ^[440]^[441]

Прогресс в области искусственного интеллекта усилил интерес к этой теме. Сторонники концепции благосостояния и прав искусственного интеллекта часто утверждают, что отрицать наличие у ИИ разума, если он появится, будет особенно легко. Они предупреждают, что это может стать моральным «слепым пятном», аналогичным рабству или промышленному животноводству, которое может привести к масштабным страданиям, если разумный ИИ будет создан и бездумно эксплуатироваться. ^[437]^[436]

Будущее

Сверхразум и сингулярность

Сверхразум — это гипотетический агент, обладающий интеллектом, значительно превосходящим интеллект самых ярких и одарённых людей. ^[426] Если исследования в области искусственного интеллекта приведут к созданию достаточно разумного программного обеспечения, оно , возможно, сможет перепрограммировать и совершенствовать себя . Улучшенное программное обеспечение будет ещё более эффективно совершенствоваться, что приведёт к тому, что И. Дж. Гуд назвал « взрывом интеллекта », а Вернор Виндж — « сингулярностью ». ^[442]

Однако технологии не могут совершенствоваться экспоненциально бесконечно и обычно следуют S-образной кривой , замедляясь, когда достигают физических пределов своих возможностей. ^[443]

Трансгуманизм

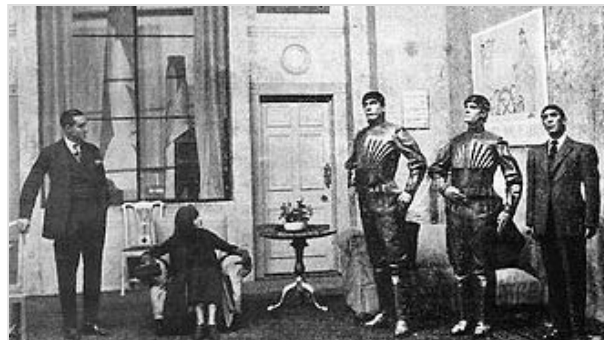
Конструктор роботов Ханс Моравец , кибернетик Кевин Уорвик и изобретатель Рэй Курцвейл предсказали, что в будущем люди и машины могут объединиться в киборгов , которые будут более способными и сильными, чем каждая из них. Эта идея, называемая трансгуманизмом, берёт своё начало в трудах Олдоса Хаксли и Роберта Эттингера . ^[444]

Эдвард Фредкин утверждает, что «искусственный интеллект — это следующий шаг в эволюции». Эта идея была впервые высказана Сэмюэлем Батлером в его книге « Дарвин среди машин » ещё в 1863 году и развита Джорджем Дайсоном в его книге 1998 года «*Дарвин среди машин: эволюция глобального интеллекта*» . ^[445]

В художественной литературе

Искусственные существа, способные мыслить, появлялись в качестве средств повествования еще со времен античности ^[446] и были постоянной темой в научной фантастике ^[447] .

Распространенный троп в этих произведениях начался с *«Франкенштейна»* Мэри Шелли, где человеческое творение становится угрозой своим хозяевам. Сюда входят такие работы, как *«Космическая одиссея 2001 года»* Артура Кларка и Стэнли Кубрика (оба 1968 года) с HAL 9000, смертоносным компьютером, управляющим космическим кораблём *«Дискавери-1»*, а также *«Терминатор»* (1984) и *«Матрица»* (1999). В отличие от них, редкие верные роботы, такие как Горт из *«Дня, когда Земля остановилась»* (1951) и Бишоп из *«Чужих»* (1986), менее заметны в массовой культуре. ^[448]



Само слово «робот» было придумано Карелом Чапком в его пьесе 1921 года *«RUR»*, название которой расшифровывается как «Россумские универсальные роботы».

Айзек Азимов ввёл Три закона робототехники во многих рассказах, наиболее ярким примером которых является сверхразумный компьютер «Мультивак». Законы Азимова часто упоминаются в неформальных дискуссиях об этике машин; ^[449] хотя почти все исследователи искусственного интеллекта знакомы с законами Азимова из популярной культуры, они, как правило, считают их бесполезными по многим причинам, одной из которых является их неоднозначность. ^[450]

В нескольких произведениях искусственный интеллект используется для того, чтобы заставить нас столкнуться с фундаментальным вопросом о том, что делает нас людьми, показывая нам искусственные существа, способные чувствовать и, следовательно, страдать. Это отражено в фильме Карела Чапека *«RUR»*, фильмах *«Искусственный интеллект»* и *«Из машины»*, а также в романе Филипа К. Дика *«Мечтают ли андроиды об электроовцах?»*. Дик рассматривает идею о том, что наше понимание человеческой субъективности меняется под воздействием технологий, созданных с помощью искусственного интеллекта. ^[451]

Смотрите также

- Искусственное сознание — область когнитивной науки
- Искусственный интеллект и выборы — использование и влияние ИИ на политические выборы
- Обнаружение контента с помощью искусственного интеллекта — программное обеспечение для обнаружения контента, созданного искусственным интеллектом
- Ассоциация по развитию искусственного интеллекта (AAAI)
- Алгоритм выбора поведения — алгоритм, который выбирает действия для интеллектуальных агентов.
- Автоматизация бизнес-процессов — Автоматизация бизнес-процессов
- Рассуждение на основе прецедентов — процесс решения новых проблем на основе решений аналогичных прошлых проблем.
- Вычислительный интеллект — способность компьютера обучаться выполнению конкретной задачи на основе данных или экспериментальных наблюдений.
- Цифровое бессмертие — Гипотетическая концепция хранения личности в цифровой форме.
- Эмерджентный алгоритм — алгоритм, демонстрирующий эмерджентное поведение

- Женский гендер в технологиях искусственного интеллекта — Гендерные предубеждения в цифровых технологиях
- Глоссарий искусственного интеллекта — список определений терминов и понятий, обычно используемых при изучении искусственного интеллекта.
- Усиление интеллекта — использование информационных технологий для усиления человеческого интеллекта.
- Интеллектуальный агент — программный агент, действующий автономно.
- Интеллектуальная автоматизация — программный процесс, сочетающий роботизированную автоматизацию процессов и искусственный интеллект.
- Список журналов по искусственному интеллекту
- Список проектов искусственного интеллекта
- Загрузка разума — гипотетический процесс цифровой эмуляции мозга
- Органоидный интеллект — использование клеток мозга и мозговых органоидов для интеллектуальных вычислений
- Роботизированная автоматизация процессов — вид технологии автоматизации бизнес-процессов
- *«Последний день»* — валлийский научно-фантастический роман 1967 года
- Wetware-компьютер — компьютер, состоящий из органического материала
- DARWIN EU — инициатива Европейского союза, координируемая Европейским агентством по лекарственным средствам (EMA) для сбора и использования фактических данных (RWE) для поддержки оценки и надзора за лекарственными средствами на всей территории ЕС.

Пояснительные записки

- a. Этот список интеллектуальных качеств основан на темах, рассматриваемых в основных учебниках по искусственному интеллекту, включая: Рассел и Норвиг (2021), Люгер и Стабблфилд (2004), Пул, Макворт и Гебель (1998) и Нильссон (1998).
- b. Этот список инструментов основан на темах, рассматриваемых в основных учебниках по искусственному интеллекту, включая: Russell & Norvig (2021), Luger & Stubblefield (2004), Poole, Mackworth & Goebel (1998) и Nilsson (1998).
- c. Это одна из причин, по которой экспертные системы оказались неэффективными для сбора знаний. ^[30] ^[31]
- d. «Рациональный агент» — общий термин, используемый в экономике, философии и теории искусственного интеллекта. Он может относиться к чему угодно, направляющему своё поведение на достижение целей, например, к человеку, животному, корпорации, государству или, в случае ИИ, к компьютерной программе.
- e. Алан Тьюринг обсуждал центральную роль обучения ещё в 1950 году в своей классической статье «Вычислительная техника и интеллект». ^[42] В 1956 году на оригинальной летней конференции по искусственному интеллекту в Дартмуте Рэй Соломонофф написал доклад о неконтролируемом вероятностном машинном обучении: «Машина индуктивного вывода». ^[43]
- f. См. AI Winter § Машинный перевод и отчет ALPAC за 1966 год.
- g. По сравнению с символической логикой, формальный байесовский вывод требует больших вычислительных затрат. Для того чтобы вывод был корректным, большинство наблюдений должны быть условно независимы друг от друга. AdSense использует байесовскую сеть с более чем 300 миллионами ребер для определения того, какие объявления показывать. ^[94]
- h. Ожидание-максимизация, один из самых популярных алгоритмов в машинном обучении, позволяет кластеризовать данные при наличии неизвестных скрытых переменных. ^[96]