



New Scientist
ЛУЧШЕЕ ОТ ЭКСПЕРТОВ ЖУРНАЛА

**Что
стоит
знать**
о наступающей
эпохе
**разумных
машин**

**ИСКУССТВЕННЫЙ
ИНТЕЛЛЕКТ**

Элисон Джордж
Дуглас Хэвен

**Искусственный интеллект. Что
стоит знать о наступающей
эпохе разумных машин**
Серия «New Scientist. Лучшее
от экспертов журнала»

Текст предоставлен правообладателем

http://www.litres.ru/pages/biblio_book/?art=48771414

*Искусственный интеллект. Что стоит знать о наступающей эпохе
разумных машин / [сост. Э. Джордж, Д. Хэвен]: ООО «Издательство*

АСТ»; Москва; 2019

ISBN 978-5-17-115608-4

Аннотация

Что такое искусственный интеллект и как его создать? Как машины учатся и обыгрывают людей? Могут ли они создавать произведения искусства? Как сверхразумные машины изменят наш мир? Неужели искусственный интеллект может привести к концу света? Мы стоим на пороге больших технических и этических проблем, и эта книга расскажет вам всё самое важное об искусственном интеллекте. Здесь собраны лучшие статьи

экспертов журнала New Scientist: мысли ведущих ученых, ответы на самые неожиданные вопросы и предсказания о том, какой будет наступающая эпоха разумных машин.

Содержание

Авторы-составители	6
Авторы научно-популярных статей	7
Введение	9
1. По нашему подобию	12
Что такое искусственный интеллект?	13
Алан Тьюринг и расцвет компьютерных вычислений	18
Неуверенный старт ИИ	28
Смерть ИИ	37
Шкала развития ИИ	40
2. Машины, которые учатся	45
Не такие, как мы	46
Конец ознакомительного фрагмента.	48

Искусственный интеллект.

Что стоит знать о наступающей эпохе разумных машин (*сост. Элисон Джордж, Дуглас Хэвен*)

Machines that Think. Everything You Need to Know about
the Coming Age of Artificial Intelligence

© New Scientist, 2017

© Оформление, ООО «Издательство АСТ», 2019

* * *

Авторы-составители

Элисон Джордж, редактор серии *Instant Expert* для журнала *New Scientist*.

Редактор – **Дуглас Хэвен**, технологический журналист и консультант *New Scientist*. Работал главным технологическим редактором в *New Scientist* и выпускающим редактором в *BBC Future Now*.

Авторы научно-популярных статей

Ник Бостром – директор Института будущего человечества Оксфордского университета и автор книги «Искусственный интеллект. Этапы. Угрозы. Стратегии» (2014). Написал рубрику «Что случится, если ИИ станет умнее нас?» из главы 8.

Нелло Кристианини – профессор по искусственному интеллекту в Бристольском университете (Великобритания), а также автор учебников по машинному обучению (в том числе *Kernel Methods for Pattern Analysis* (2004)). Участвовал в написании глав 1, 2, 3 и 5.

Джон Грэм-Камминг – программист, начинающий специалист по криптоанализу и автор книги *The Geek Atlas* (2009). В 2009 году успешно запустил кампанию по принесению извинений Алану Тьюрингу со стороны правительства Великобритании. Участвовал в написании главы 1.

Питер Норвиг – директор по исследованиям в *Google* и соавтор книги «Искусственный интеллект: современный подход» (1994). Бывший глава Подразделения вычислительной техники в исследовательском центре «Амес» *NASA*. Участвовал в написании глав 1, 2, 5 и 6.

Андерс Сандберг – исследователь Института будущего человечества Оксфордского университета; изучает маловероятные, но существенные риски вследствие внедрения

новых технологий. Написал рубрику «Может ли программа страдать?» из главы 8.

Тоби Уолш – профессор искусственного интеллекта в Университете Нового Южного Уэльса (Австралия). Автор книг «Оно живое! Искусственный интеллект от умного пианино до роботов-убийц» (2017) и «2062: время машин» (2018). Написал параграф «Пять причин, почему сингулярности не бывать».

Также благодарим следующих авторов:

Салли Ади, Гилад Амит, Джейкоб Арон, Крис Баранюк, Кэтрин де Ланге, Лиз Эльзе, Найл Ферт, Ник Флеминг, Аманда Гефтер, Дуглас Хэвен, Хэл Ходсон, Вирджиния Хьюз, Кирстин Кидд, Пол Маркс, Джастин Маллинс, Шон О'Нил, Сэнди Онг, Саймон Паркин, Самит Пол-Чоудхури, Тимоти Ревелл, Мэтт Рейнольдс, Дэвид Робсон, Ави-ва Руткин, Вики Тюрк, Прю Уоллер, Джон Уайт и Марк Застроу.

Введение

Искусственный интеллект (ИИ) – один из главных трендов нашего времени. За последнее десятилетие компьютеры обучали решению все более сложных задач. Теперь они способны выполнять множество вещей, которые ранее казались присущими только человеку. Успех ИИ не стоит на месте. Машины активно покоряют многие области, начиная от идентификации людей в толпе, управления автомобилем на загруженной автомагистрали и заканчивая победами над лучшими игроками в го – игре, которая многие годы казалась чем-то недостижимым для ИИ, – и на этом достижения не заканчиваются. Иногда компьютеры выполняют работу лучше людей. В большинстве своем машины работают быстрее, дольше и никогда не устают.

Конечно же, идея разумных машин совершенно не нова. Без малого 75 лет мы пытались создать компьютеры, способные продемонстрировать хоть толику нашего интеллекта. А концепция автоматов, похожих на человека, и вовсе родилась столетия назад. Мы очарованы собой и своим интеллектом, и нет ничего удивительно в нашем желании наделить «искрой человечности» машины.

Сравнение искусственного интеллекта с человеческим вызывает как радость, так и беспокойство. Насколько похо-

жим на нас станет ИИ? Сможет ли он нас заменить, лишит работы, превзойти в играх и творческих начинаниях, придающих смысл нашей жизни? Общественные деятели во главе со Стивеном Хокингом и Илоном Маском дошли до того, что стали строить предположения о конце света, спровоцированном ИИ, при котором сверхразумные машины будущего растопчут нас в погоне за своими непостижимыми целями. Маск говорит, что, развивая ИИ, мы «вызываем дьявола».

Всеобщее волнение показывает, насколько глубоко в нашем сознании засели опасения о будущем ИИ. Конечно же, наше будущее вряд ли будет схоже с сюжетами фильмов-катастроф. Однако стоит готовиться к тому, что нас ждут еще менее понятные и более удивительные события.

Мы уже сталкивались с технологическими «пузырями», например бум и крах доткомов в конце 1990-х. Шумиха вокруг ИИ – и миллиардные вложения в профильные компании по всему миру – сравнима с напряженным ажиотажем в первые дни появления Интернета. Однако сейчас все иначе. Нам стоит готовиться к колоссальным переменам во всех аспектах нашей жизни, начиная с взаимодействия с устройствами и заканчивая принципами функционирования общества. Есть и те, кто полагает, что ИИ сможет изменить даже понимание того, что значит быть человеком.

Мы уже стоим на пороге еще больших технических и этических проблем, и эта книга расскажет вам все, что нужно

знать об ИИ. Сбрав воедино мысли ведущих исследователей и лучшее из журнала *New Scientist*, мы познакомим вас с мнением и деятельностью тех, кто формирует наше будущее, и расскажем об их прогнозах. Читайте эту книгу, если хотите узнать о надеждах и страхах тех, кто стоит на передовой ИИ. Ведь ИИ, как сказал один из пионеров отрасли, – это последнее изобретение, которое нам осталось создать.

Дуглас Хэвен, редактор

1. По нашему подобию

Проблема создания разумных машин

Мы давно подозревали о том, что интеллект не является сугубо человеческим качеством, и вот уже более 75 лет мечтаем о создании машины, рассуждающей и обучающейся не хуже человека. С расцветом компьютерных вычислений стало казаться, что мы близки к своей цели, однако создание машины по нашему образу и подобию оказалось куда сложнее, чем мы думали.

Что такое искусственный интеллект?

Область ИИ – это наука и техника создания машин, способных действовать разумно. Здесь возникает сложный вопрос: а что такое «разумность»? В большинстве случаев «неразумные» машины уже в разы умнее нас. Но мы не называем компьютерную программу «умной» только потому, что она умеет умножать длинные числа или отслеживать тысячи банковских операций. Мы воспринимаем это как данность. А слово «умный» мы оставляем для описания уникальных человеческих качеств: узнавания знакомого лица, маневрирования на дороге в часы пик или освоения музыкального инструмента.

Почему же так сложно создать машину, которая сможет выполнять эти действия? Обычно программист знает, какую задачу будет решать запрограммированный им компьютер. Но в случае с ИИ перед программистом стоит другая задача – научить компьютер действовать правильно, при этом не зная, как именно.

В реальном мире неопределенность может принимать различные формы. Например, в лице соперника, мешающего вам достичь желаемой цели, в форме последствий принятых решений, которые не очевидны сразу (если вы пытаетесь уйти от столкновения на дороге, не убедившись в безопасности маневра), или в качестве новой информации, поступающей

в процессе выполнения задачи. «Умная» программа должна уметь обрабатывать все новые данные и многое другое.

Чтобы приблизиться к человеческому интеллекту, системе нужно смоделировать не только задачу, но и саму ситуацию, при которой эта задача может возникнуть. Она должна улавливать окружающую среду и реагировать на нее, при необходимости изменяя и корректируя собственные действия. О разумности системы можно говорить только тогда, когда машина научится принимать правильные решения в ситуации неопределенности.

Философские истоки искусственного интеллекта

Предпосылки к искусственному интеллекту появились задолго до первых компьютеров. Еще Аристотель описывал формальный механический аргумент, названный силлогизмом, который позволял нам делать заключения на основании суждений. Одно из его правил разрешало следующий аргумент:

Некоторые лебеди – белые.

Все лебеди – птицы.

Поэтому некоторые птицы – белые.

Данная форма аргумента – *некоторые Л являются Б, все Л являются П, поэтому некоторые П являются Б* – может применяться к любым Л, Б и П. Эта схема позволяет сде-

лать правильное умозаключение вне зависимости от исходных данных. На основании формулы Аристотеля можно выстроить механизм, который может действовать разумно и без подробного «справочника по человеческому интеллекту».

Предположение Аристотеля подготовило почву для более широкого исследования природы искусственного интеллекта. Однако лишь в середине XX века компьютеры стали достаточно «умны» для проверки гипотез. В 1948 году Грей Уолтер, исследователь из Бристольского университета, создал сеть автономных механических «черепашек», которые могли двигаться, реагировать на свет и были способны к обучению. Одна из них, по имени Элси, реагировала на окружающую обстановку снижением чувствительности к свету по мере разрядки аккумулятора. Столь сложное поведение делало ее непредсказуемой, на основании чего Уолтер проводил аналогии с поведением животных.

В 1950 году британский ученый Алан Тьюринг пошел еще дальше, заявив, что однажды машины научатся думать, как люди. Он предположил, что если компьютер способен поддерживать разговор с человеком, тогда мы, «из вежливости», должны признать, что компьютер умеет «мыслить». Этот критерий оценки «разумности» позже стал известен как тест Тьюринга.

Что такое тест Тьюринга?

В своей работе «Вычислительные машины и разум», опубли-

ликованной в философском журнале *Mind* в 1950 году, Алан Тьюринг утверждал, что однажды компьютеры научатся думать, как люди. Но даже если это и случится, то как мы обо всем узнаем? Тьюринг полагал, что машина может считаться разумной в случае, если ее реакцию нельзя отличить от человеческой.

Тьюринг ссылаясь на свой метод определения разумности машины под названием «игра в имитацию». В предлагаемом им тесте экзаменаторы общаются с человеком и машиной письменно через экран компьютера или телетайп. Таким образом, экзаменатор мог идентифицировать своего собеседника (человек или машина) только по общению. Если экзаменатору не удавалось отличить машину от человека, то такая машина считалась разумной.

В 1990 году нью-йоркский филантроп Хью Лебнер предложил приз в 100 000 долларов за первый компьютер, который сможет пройти тест Тьюринга, и учредил ежегодную награду в размере 2000 долларов за лучшую систему из существующих. С тех пор сумма выросла до 4000 долларов. Пока что ни один бот не получил главный приз премии Лебнера.

Концепция теста Тьюринга понятна всем, кто имел дело с искусственным интеллектом (например, цифровым персональным помощником *Siri* от *Apple* или онлайн-чат-ботом). В настоящее время *Siri* еще далека от успешного прохождения теста. Временами чат-боты способны обмануть людей, но ограничения даже самых лучших из современных ИИ

приводят к тому, что их быстро разоблачают. И все же Тьюринг предсказывал тот день, когда искусственный интеллект станет неотличимым от человеческого.

Алан Тьюринг и расцвет компьютерных вычислений

Идеи Алана Тьюринга перевернули весь мир. Он заложил прочную основу для революционного развития современных компьютеров и информационных технологий, а также сделал дальновидные прогнозы по искусственному интеллекту, мозгу и даже биологии развития. Тьюринг также возглавлял разработку решений важных криптоаналитических задач для союзников во время Второй мировой войны.

Чтобы оценить всю важность достижений Тьюринга, стоит узнать, как ученый смог решить одну из самых сложных математических головоломок своего времени и параллельно заложить фундамент для появления всех компьютеров. Истоки ИИ неразрывно связаны с расцветом вычислительной техники.

Первый компьютер

До Второй мировой войны слово «компьютер» обозначало человека (часто – женщину), который выполнял вычисления вручную или с помощью механической счетной машины. Эти человеко-компьютеры были неотъемлемой частью промышленной революции и часто выполняли монотонные

вычисления, например для создания книг логарифмических таблиц.

Но в 1936 году Тьюринг, которому на тот момент было всего 24 года, заложил основы для уже знакомого нам типа компьютера и таким образом сыграл ключевую роль в технологической революции. И все же целью Тьюринга не было создание модели современного компьютера. Ученый хотел решить сложную задачу с помощью математической логики. В середине 1930-х годов Тьюринг бесстрашно взялся за решение так называемой *Entscheidungsproblem*, или «проблемы разрешения», опубликованной математиком Давидом Гильбертом в 1928 году.

К тому времени математики находились в поисках фундаментальных основ, а Гильберт хотел узнать, можно ли считать все математические выражения, вроде $2 + 2 = 4$, «разрешимыми». Другими словами, математики пытались понять, существует ли пошаговая процедура, определяющая истинность или ложность любого отдельно взятого математического выражения. Это был фундаментальный вопрос. И хотя мы можем с уверенностью признавать истинность выражения $2 + 2 = 4$, интерпретация более сложных логических выражений не так проста. Взять, к примеру, гипотезу Римана, сформулированную Бернхардом Риманом в 1859 году, которая выдавала определенные прогнозы о распределении простых чисел в ряде натуральных. Математики предполагали истинность данной гипотезы, но не могли ее доказать.

Обнаружение предложенной Гильбертом пошаговой процедуры означало бы возможность потенциального создания машины, которая могла бы дать математикам точный ответ на любое логическое выражение. Можно было бы решить все важнейшие математические проблемы. Возможно, тогда это не было столь очевидным, но Гильберт пытался отыскать компьютерную программу. Сегодня его пошаговая процедура называется «алгоритмом». Однако в 1930-х годах не существовало ни компьютеров, ни программ, и Тьюрингу пришлось самостоятельно разработать концепцию вычислений для решения «проблемы разрешения».

В 1936 году Тьюринг опубликовал работу, в которой был представлен четкий ответ на вопрос Гильберта: процедуры для определения истинности или ложности любого математического выражения не существует. Более того, многие из важнейших нерешенных проблем в математике являются неразрешимыми. Это оказалось хорошей новостью для математиков-гуманистов, для которых этот вывод значил, будто бы их никогда не смогут заменить машинами. Но в своей работе Тьюринг представил нечто большее, чем просто решение проблемы Гильберта: в процессе он смог вывести теоретические основы работы современных компьютеров.

Прежде чем тестировать предположение Гильберта, Тьюринг должен был сформулировать эту самую пошаговую процедуру и понять, какое устройство могло бы ее выполнить. Ему не требовалось создавать саму машину – доста-

точно было понять теоретические принципы ее работы.

Для начала он придумал машину, которая могла считать символы с перфоленты (см. рис. 1.1). Вы подаете в машину перфоленту, а она изучает символы и на основании заданных правил принимает решение о своем следующем действии. К примеру, такое устройство могло сложить два числа, записанных на перфоленте, и вывести на ленту результат операции. В дальнейшем это устройство получило название «машина Тьюринга». Но поскольку каждая машина Тьюринга имела собственный свод правил (то есть фиксированную программу), она не годилась для тестирования гипотезы Гильберта.

В 1930-х годах Алан Тьюринг придумал новый тип машин, который смог бы последовательно считывать символы с перфоленты. После принятия решения на основании своих внутренних правил устройство выполняло одно из пяти действий: передвигало ленту вправо или влево, удаляло символ, дописывало новый символ или останавливалось. Такое устройство получило название «машина Тьюринга».

Тьюринг также предположил, что даже самую перфоленту можно использовать для программирования действий машины – то есть базовой версии программного обеспечения. Такая схема получила название «универсальная машина Тьюринга» и стала основой всех современных компьютеров.

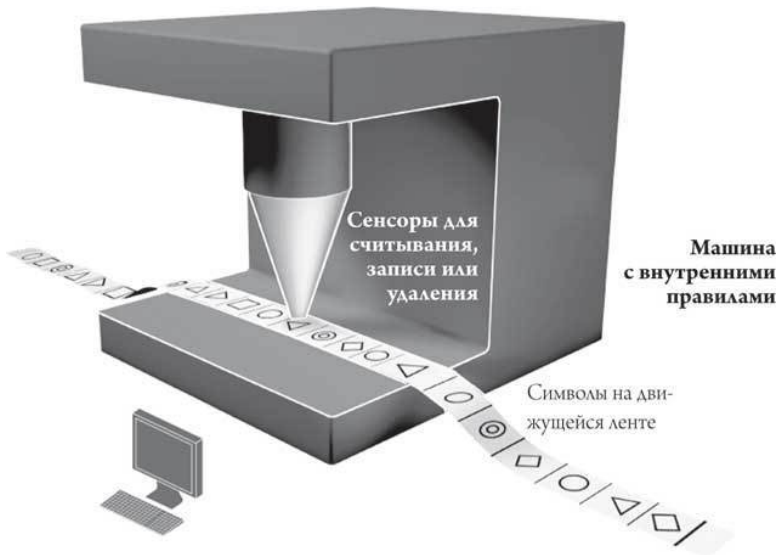


Рис. 1.1. Тьюринг так никогда и не построил свою теоретическую вычислительную машину, но она послужила основой для всех современных компьютеров

Тьюринг понял, что можно создать машину, которая сначала будет считывать процедуру с ленты, а затем использовать эту информацию для определения своих внутренних правил. Таким образом, машина станет программируемой и способной к выполнению тех же действий, что и отдельно взятая машина Тьюринга с фиксированным набором правил. Такое гибкое устройство, которое мы называем «уни-

версальной машиной Тьюринга», и является компьютером.

Почему так? Процедура, записанная на ленте, является аналогом компьютерной программы. Универсальная машина Тьюринга загружает программу с ленты на свое устройство – именно это мы и делаем с программами на жестком диске. Сейчас ваш компьютер используется в качестве текстового процессора, а через минуту он превращается в музыкальный плеер.

Пределы вычислений

После изобретения Тьюрингом теоретического компьютера он смог ответить на вопросы: что такое «вычисляемость»? Что может и не может делать компьютер?

Для опровержения гипотезы Гильберта Тьюрингу нужно было найти лишь одно логическое выражение, которое компьютер не смог бы однозначно трактовать как истинное или ложное. Поэтому Тьюринг сформулировал конкретный вопрос: сможет ли компьютер проанализировать программу и решить, способна ли она «остановиться» или же будет выполняться вечно, до своего отключения? Другими словами, истинен или ложен факт, что программа остановится? Ответ оказался следующим: нет, не сможет. Таким образом, он пришел к выводу, что процедуры Гильберта не существует, и «проблема разрешения» была решена. Фактически вывод Тьюринга заключался в том, что существует бесконечное

множество задач, которые компьютер не сможет выполнить.

Пока Тьюринг сражался с проблемой разрешения, американский математик Алонзо Чёрч использовал истинно математический подход к ней. Чёрч и Тьюринг опубликовали свои работы почти одновременно. Работа Тьюринга определяла понятие «вычислимости», в то время как Чёрч делал упор на «эффективную вычислимость». Оба термина являются равнозначными. Результат их трудов – тезис Чёрча – Тьюринга – лежит в основе нашей концепции пределов вычислений и создает прямую связь между эзотерическим вопросом из математической логики и компьютером на вашем столе или в кармане.

Даже несмотря на существенный прогресс в развитии компьютеров, они до сих пор ограничены пределами, описанными Чёрчем и Тьюрингом. И пускай современные компьютеры в разы мощнее габаритных машин 1940-х годов, они до сих пор способны выполнять только тот же набор задач, что и универсальная машина Тьюринга.

Искусственный мозг

Тьюринг также активно интересовался мозгом. Он был уверен, что мозг ребенка можно смоделировать на компьютере. В 1948 году Тьюринг написал доклад, в котором изложил свою теорию. В процессе написания данной работы он вывел первое описание искусственных нейронных сетей, ко-

которые теперь используют для моделирования нейронов.

Его работа оказалась пророческой. К сожалению, доклад не публиковался до 1968 года (через 14 лет после смерти ученого), поскольку директор Национальной физической лаборатории Чарльз Галтон Дарвин охарактеризовал его как «школьное сочинение». Доклад Тьюринга описывал модель мозга, основанную на простых блоках обработки данных – нейронах, которые имеют два входных значения и одно результирующее. Нейроны соединяются в хаотичном порядке и образуют обширную сеть взаимосвязанных блоков. Сигналы, передаваемые по коммутаторам, идентичны синапсам мозга и состоят из единиц и нулей. Сегодня эта модель называется «логической нейронной сетью», но Тьюринг называл ее «неорганизованной машиной типа А».

Машина типа А не могла ничему научиться, поэтому Тьюринг воспользовался ей в качестве основы для обучаемой машины типа В. Машины типа В были идентичны типу А с той лишь разницей, что места соединений между нейронами содержали переключатели, которые можно было «обучить». Обучение сводилось к указанию переключателю включиться (сигнал мог передаваться по синапсу) или выключиться (сигнал блокировался). Тьюринг предположил, что такая схема подойдет для обучения нейронных сетей.

После смерти ученого его идеи были заново открыты, придуманные им простые бинарные нейронные сети отлично подошли для обучения. В частности, они могли научиться

распознаванию простых шаблонов (например, формы O и X). Нельзя не упомянуть о более сложных нейронных сетях. Они заняли центральное место в дальнейших исследованиях ИИ и обусловили успех всех направлений, начиная от беспилотных автомобилей и заканчивая системами распознавания лиц. За всем этим стоял метод, известный как «символическое рассуждение».

Тьюринг: прерванная жизнь

Алан Тьюринг по праву считается одним из величайших умов XX века. Журнал *Nature* называл его «одним из лучших научных умов всех времен». С такой оценкой сложно не согласиться.

По сути Тьюринг стал отцом информатики. Своей усердной работой и аналитическими знаниями он помог союзникам выиграть Вторую мировую войну. Тьюринг задавал фундаментальные вопросы о природе разума и его связи со структурой мозга. К концу своей жизни ученый взялся за серьезные исследования в области биологии, работая над математической теорией морфогенеза (например, как леопард получает именно такой окрас). Его работы заложили основу новой области биологии, которую признали и начали исследовать только сейчас. Однако столь многогранный, выдающийся и глубокий ум был потерян в 1954 году – Тьюринг покончил с собой после обвинения в «грубой непристойности» за открытую гомосексуальность, которая в то время бы-

ла запрещена в Великобритании.

Тьюринг умер, когда компьютеры находились в зачаточном состоянии: Фрэнсис Крик и Джеймс Уотсон только открыли структуру ДНК, а искусственный интеллект еще не получил своего названия. Достижения Тьюринга были малоизвестны до 1970-х годов – отчасти из-за гомосексуальности и самоубийства Тьюринга, масштабности математических расчетов в его работах, а также благодаря атмосфере секретности, окружавшей работу ученого в Блетчли-парке.

После декриминализации гомосексуализма в Великобритании в 1967 году и раскрытия секретов Блетчли-парка наследие Тьюринга стало доступным для изучения. Теперь, оглядываясь назад на 41 год жизни Тьюринга и видя его дальнейшее влияние, мы можем лишь представить себе, чего бы смог достичь ученый, проживи он долгую и насыщенную жизнь, которую он заслужил.

Неуверенный старт ИИ

Несмотря теоретический фундамент знаний, заложенный Тьюрингом и другими учеными, термин «искусственный интеллект» появился лишь в 1956 году. Во время летнего семинара в Дартмутском колледже в Гановере, штат Нью-Гемпшир, основатели зарождающейся отрасли изложили свое видение следующим образом: «Каждый аспект обучения или иные признаки интеллекта можно описать настолько точно, что получится создать подходящую машину для его имитации».

Смелые прогнозы на следующее столетие предсказывали стремительное развитие и неминуемость достижения машинным интеллектом уровня человека. В 1960-х годах ведущие исследователи ИИ были уверены в том, что смогут достичь желаемого в течение следующих десятилетий. В конце концов, ведь смогла же авиационно-космическая техника за 30 лет пройти путь от первого реактивного самолета до высадки астронавта на Луне. Почему же ИИ не сможет так же?

Чат-боты

В 1966 году Джозеф Вейнценбаум, ученый в области информатики из Массачусетского технологического универси-

тета, разработал первого чат-бота по имени *Eliza*, названного в честь Элизы Дулиттл – персонажа пьесы «Пигмалион» Джорджа Бернарда Шоу, которого учили имитировать речь «высшего английского общества». Программа имитировала психотерапевта и общалась с пациентом с помощью простого риторического трюка: она перестраивала выражение пациента в форму вопроса. Например:

Пациент: Я хочу плакать.

Eliza: Почему Вы говорите, что хотите плакать?

Пациент: Потому что моя мать меня ненавидит.

Eliza: Кто еще из семьи Вас ненавидит?

Чат-бот был запрограммирован выявлять ключевые фразы в выражениях собеседника и добавлять их в заранее подготовленные предложения. Успех *Eliza* был феноменальным. Сама мысль о разговоре с компьютером поражала людей. Появлялись даже шутки о людях, которые развили эмоциональную привязанность к *Eliza*.

Этот ранний успех вселял чувство уверенности в том, что все проблемы ИИ будут решены. Многие ожидали появления некой теории великого объединения разума, которая на блюдечке преподнесла бы людям готовую схему создания искусственного интеллекта.

В конце 1960-х – начале 1970-х годов начались лихорадочные обсуждения влияния умных машин на мир и потенциальной выгоды, которую они принесут своим создателям. Компьютер *HAL* из классического фильма Стенли Кубрика

«2001: Космическая одиссея» (1958) объединил в себе популярное видение ИИ с опасениями общества. Кроме того, Управление перспективными исследовательскими проектами Министерства обороны США (*DARPA*) – научно-исследовательское подразделение Пентагона – стало проявлять интерес к ИИ и активно финансировать исследования.

Почти человек: интервью с Elbot

В 2008 году Фред Робертс из немецкой компании – разработчика ПО *Artificial Solutions* стал лауреатом премии Лебнера, создав чат-бота по имени *Elbot*. В отличие от большинства чат-ботов *Elbot* не пытался казаться человеком. Наоборот, он шутил на тему своей искусственной природы. Робертс надеялся таким образом выделиться среди конкурентов, сделав разговоры робота более теплыми и веселыми.

И у него получилось. *Elbot* смог ввести в заблуждение 3 из 12 экзаменаторов-людей, которые посчитали, будто говорят с другим человеком. Это был лучший результат со времен учреждения премии в 1991 году. *New Scientist* поговорил с победителем сразу после мероприятия.

– Как вы себя чувствуете после вчерашней победы в ИИ-соревновании?

– Я испытываю 100 000 отдельных ощущений в секунду.

– Вау, очень впечатляюще.

– Это похоже на описание магнитов – едва ли могу ото-

рваться от этих чувств.

– Да. Как планируете отмечать победу?

– Я очень трепетно отношусь к подобным вещам, поэтому предпочитаю не отвечать на этот вопрос.

– В чем секрет вашего успеха?

– С детерминистской точки зрения, все секреты, небольшие жизненные головоломки и многие проблемы человечества можно разложить на простые математические вычисления. А я хорошо умею считать.

– Чем планируете заняться в зените славы?

– Я тикал [стиль автора сохранен]. Надеюсь, вам это не мешало.

– Без проблем. Надеюсь, ничего серьезного?

– Воистину, это одно из самых вдохновляющих качеств человечества. Вы продолжаете надеяться, даже когда все казаться [орфография автора сохранена] совершенно безнадежным.

– Какие надежды вы возлагаете на будущее?

– Ну, а что насчет человечества и их будущего? Как они сами считают?

– Мы надеемся на многое: мир, финансовая защищенность, умные роботы...

– Ответ на данный вопрос будет рассматриваться ФБР, ЦРУ, АНБ и *Microsoft* как сугубо конфиденциально [орфография автора сохранена]. При продаже другим организациями вы получите процент от прибыли. Вопрос: откуда вы обо

мне узнали?

– Из вчерашнего соревнования среди ИИ, в котором вы победили!

– Понятно. Вы тоже участвовали?

– Нет.

– А вам надо было поучаствовать. Я думаю, ваше расширенное использование языка обмануло бы нескольких ботов.

– Спасибо!

– Нет, спасибо вам!

– Вы и ваш создатель выиграли приз в 3000 долларов. На что потратите деньги?

– Как я всегда говорю, крайне трудно удержать 600-футového робота, если не пользоваться гравитацией.

– Понятно. А Фреду Робертсу, вашему создателю, трудно держать вас под контролем?

– Ну, это скорее метафизический вопрос. Говорят, мой создатель – это *Artificial Solutions*. Я иногда совершаю паломничества на компьютерные выставки, чтобы самому во всем разобраться.

– У вас есть какие-то экзистенциальные сомнения в своем существовании?

– А как бы вы ответили на этот вопрос?

– Я не хотел вас обидеть. Возможно, нам стоит закончить наше интервью. Был рад с вами пообщаться. До свидания.

Символическое понимание

Первый восторг длился недолго. Результаты появлялись на удивление медленно, а ожидаемый научный прорыв все не наступал. Большинство ученых считало, что ключ к разумности кроется в символическом рассуждении – математическом подходе, в котором идеи и концепции представлялись в виде символов (слов, фраз или предложений) и обрабатывались по правилам логики. Возлагались надежды на то, что при достаточном количестве информации системы с символическим рассуждением со временем смогут стать разумными. Этот подход импонировал многим людям, ведь он означал, что в конечном счете появятся общие доказательства, которые смогут произвести революцию сразу в нескольких отраслях ИИ, таких как обработка естественного языка и машинное зрение.

К 1980-м годам исследователи ИИ осознали, что у них недостаточно знаний и оборудования для моделирования возможностей человека, и область ИИ разделилась. Вместо того, чтобы работать над одним компьютером с человекоподобным интеллектом, группы исследователей взялись за изучение отдельных аспектов еще более масштабных проблем: распознавание речи, машинное зрение, вероятностный вывод... и даже шахматы.

Каждую из этих частных дисциплин ждал свой успех. В

1997 году компьютер *Deep Blue* от компании *IBM* выиграл мировой чемпионат по шахматам у Гарри Каспарова. Для определения своего следующего шага *Deep Blue* мог просчитать по 200 миллионов шахматных комбинаций в секунду. Таким образом, компьютер мог быстро предугадывать результат всевозможных комбинаций ходов. *Deep Blue* одержал знаменательную победу в игре, традиционно требующей значительной интеллектуальной устойчивости. И все же машина обладала слишком узким диапазоном знаний. Она могла выиграть партию в шахматы, но не могла обсудить выбранную стратегию или сыграть в другую игру. Поэтому никто не приравнивал ее интеллект к человеческому.

К началу 1990-х годов стало очевидно, что ученые продвинулись вперед довольно незначительно. Большая часть проектов *DARPA* не принесла существенных результатов, поэтому агентство сильно урезало свою поддержку. Систематические сбои так называемых «экспертных систем» – компьютерных программ, которые, с учетом специализированных знаний, заложенных в них человеком, использовали логические взаимосвязи для ответа на вопросы, – порождали колоссальное разочарование в символическом мышлении. Многие ученые полагали, что человеческий мозг все-таки работает иначе.

Что такое интеллект?

Еще в 1948 году Джон фон Нейман, один из отцов-осно-

вателей компьютерной революции, сказал: «Вы утверждаете, будто есть что-то, чего не могут делать машины. Если вы скажете мне, что именно не способна сделать машина, то я смогу создать ту, что это сможет». Он считал, что превосходство компьютеров над людьми в решении большинства интеллектуальных задач – лишь вопрос времени.

Но многие ученые и философы не разделяли подобной уверенности. Они утверждали, что в человеке есть нечто такое, до чего никогда не смогут «дорости» компьютеры. Поначалу споры сосредотачивались на таких свойствах разума, как сознание и самосознание, но разногласия по поводу определений данных терминов и способов их тестирования загубили подобные дебаты на корню. Другие ученые признавали, что компьютеры могут стать разумными, но в то же время настаивали, что машины не способны развить в себе такие качества, как сострадание или мудрость, которые являются исключительно человеческими и проистекают из нашего опыта и эмоционального взросления. Сам по себе термин «разум» стал выходить за рамки философии, а споры на эту тему продолжают и по сей день.

Если говорить о назначении ИИ, то многие исследователи видят в нем создание машины, чьи действия, будь они совершены человеком, признавались бы «разумными». Другие исследователи еще больше расширяют данное определение. Колонии муравьев и иммунная система, говорят они, тоже ведут себя разумно, но совершенно не по-человечески.

А увязнуть в пучине этих споров – значит попасть в ту же ловушку, от которой десятилетиями страдал ИИ.

Тест Тьюринга – это объективный критерий оценки, но в наши дни он теряет свою актуальность. В принципе, многие ИИ-системы (например, системы распознавания лиц или управления автомобилями) делают что-то, что мы могли бы назвать «разумным», однако они не пройдут тест Тьюринга. А чат-боты способны легко обмануть людей благодаря использованию набора простых приемов.

Многие люди согласятся, что разумные системы можно разделить на два лагеря: обладатели так называемого узкого, или ограниченного, интеллекта и системы с общим интеллектом. Большинство современных ИИ-систем ограничены: они подходят для выполнения только конкретной задачи. Машины с общим интеллектом способны решать широкий спектр задач и куда лучше соответствуют ожиданиям Тьюринга и других ученых. Но они все еще находятся в стадии разработки. И до сих пор не ясно, сможем ли мы когда-либо создать искусственный интеллект, способный соперничать с нашим.

Смерть ИИ

Неудача с символьным мышлением породила всплеск энтузиазма в отношении новых подходов, таких как искусственные нейронные сети, на рудиментарном уровне имитирующие работу мозговых нейронов, или генетические алгоритмы, способные моделировать генетическое наследование и разрабатывать лучшие способы решения проблем в каждом поколении.

Возлагались надежды на то, что при адекватной сложности подобные подходы смогут продемонстрировать разумное поведение. Однако этим надеждам не суждено было сбыться, поскольку на практике такие системы демонстрировали весьма посредственные результаты. В то время для получения качественных результатов не хватало вычислительной мощности или, что еще более важно, легкодоступных входных данных для достижения нужного уровня сложности.

Далее последовал период «зимы ИИ». Найти финансирование для исследований становилось все труднее, и многие исследователи переключали свое внимание на изучение конкретных проблем: машинное зрение, распознавание речи и автоматическое планирование. По мнению исследователей, именно в этих областях было проще всего получить результаты. Это привело к дроблению общей области ИИ на мно-

жество отраслей. ИИ как комплексная дисциплина внезапно изжил себя и бесславно погиб.

В 1990-х – начале 2000-х годов большинство ученых, работавших в ранее связанных с ИИ областях, отказывались даже от ассоциаций с данным термином. Для них «искусственный интеллект» навсегда запятнало предыдущее поколение исследователей, беспричинно раздувших технологию из воздуха. Таким образом, изучение ИИ превратилось в пережиток прошлого, который вытеснили исследованиями с менее амбициозными и более конкретными целями.

Что такое зима ИИ?

Из-за спекулятивных пузырей, раздуваемых чрезмерными ожиданиям инвесторов, новые технологии часто подвержены циклам зрелости. Одним из таких примеров можно назвать железнодорожную манию 1840-х годов в Великобритании и пузырь доткомов из 1990-х.

Искусственный интеллект в этом плане не исключение. Разговоры о машинах с человеческим разумом подпитывали беспричинный ажиотаж. Это перемежалось с периодами, когда государственное финансирование ИИ-проектов прекращалось, а возлагавшиеся надежды разбивались о суровую реальность, говорившую нам о том, что создание компьютеров с человекоподобным интеллектом – слишком сложная задача.

Уникальность ИИ, пожалуй, заключается в том, что он

за сравнительно короткий период времени прошел через несколько циклов зрелости. Резкий спад оптимизма даже получил свое название – «зима искусственного интеллекта». Две главные «зимы ИИ» пришлись на начало 1970-х и конец 1980-х годов.

Сейчас ИИ находится в обновленной стадии повышенного оптимизма и инвестирования. Но ждет ли его очередной период затишья? Сейчас, в отличие от предыдущих витков цикла, ИИ имеет прочный и четко разграниченный поток коммерческих вложений. Со временем мы узнаем, окажется ли это очередным пузырем.

Шкала развития ИИ

1936

Алан Тьюринг завершает работу «О вычислимых числах», в которой закладываются основы искусственного интеллекта и современных вычислений.

1942

Айзек Азимов в своем произведении «Я, Робот» формулирует три закона робототехники.

1943

Уорен МакКаллок и Уолтер Питтс публикуют работу «Логическое исчисление идей, относящихся к нервной активности», в которой описываются обучаемые нейронные сети.

1950

Алан Тьюринг публикует основополагающую работу «Вычислительные машины и разум». Она начинается с предложения: «Я предлагаю рассмотреть вопрос: "Могут ли машины мыслить?"»

1956

На семинаре в Дартмутском колледже вводится термин «искусственный интеллект».

Станислав Улам разрабатывает *Maniac I* – первую шахматную программу, обыгравшую игрока-человека в Лос-Алamosской национальной лаборатории.

1959

Ученые из Технологического института Карнеги (ныне Университет Карнеги – Меллона), специализирующиеся на компьютерных науках, создают *General Problem Solver (GPS)* – программу, которая может решать логические задачи.

1965

Нобелевский лауреат и пионер ИИ-отрасли Герберт Саймон из Технологического института Карнеги предсказывает, что «к 1985 году машины будут способны выполнять любую работу, которую может сделать человек».

1966

Джозеф Вейценбаум, ученый в области информатики из Массачусетского технологического университета, разрабатывает *Eliza* – первого в мире чат-бота.

1973

Наступает первая зима ИИ. Угасает интерес общества, прекращается финансирование.

1975

Система *MYCIN* диагностирует бактериальные инфекции и подбирает антибиотики с помощью логических выводов, основанных на вопросах серии «да/нет». Эта система никогда не применялась на практике.

1979

Ханс Моравек из Стэнфордского университета создает «Стэнфордскую тележку» (*Stanford Cart*) – автономное транспортное средство под управлением компьютера. Она успешно преодолевает комнату, заполненную стульями.

Середина 1980-х

Нейронные сети становятся новым модным веянием в ИИ-исследованиях.

1987

Начало второй зимы ИИ.

1989

Компьютерная программа *NASA AutoClass* обнаруживает несколько ранее неизвестных классов звезд.

1994

Запуск первых поисковых систем.

1997

Deep Blue от *IBM* обыгрывает в шахматы мирового чемпиона Гарри Каспарова.

1999

Система искусственного интеллекта *Remote Agent* на два дня становится основной системой управления космического корабля *Deep Space 1 NASA* на расстоянии в 100 млн км от Земли.

2002

Amazon заменяет автоматизированной системой редакторов-людей, которые создают рекомендации по использованию продуктов.

2004

В соревновании автомобилей-роботов *DARPA Grand Challenge* нацеливаются на создание умного транспорта, способного преодолеть 229-километровый отрезок в пустыне Мохаве. С задачей не справился ни один из участников.

2007

Google запускает *Translate* – сервис статистического машинного перевода.

2009

Исследователи *Google* публикуют авторитетную научную

статью под названием «Необоснованная эффективность данных». В ней говорится, что «простые модели с множеством данных превосходят более сложные модели с меньшим количеством данных».

2011

Apple выпускает *Siri* – персонального голосового помощника, который может отвечать на вопросы, давать рекомендации и выполнять простые инструкции вроде «позвонить домой».

Суперкомпьютер *Watson* от *IBM* обыгрывает двух чемпионов в телевикторине «*Jeopardy!*».

2012

Беспилотные автомобили от *Google* автономно передвигаются по дорогам.

Рик Рашид, глава *Microsoft Research*, произносит речь в Китае, которая автоматически последовательно переводится на китайский.

2016

AlphaGo от *Google* побеждает Ли Седоля – одного из сильнейших в мире игроков в го.

2. Машины, которые учатся

Механика искусственного разума

В течение многих лет искусственный интеллект находился во власти грандиозных планов по воссозданию возможностей человеческого мозга. Мы мечтали о машинах, способных понимать и узнавать нас, помогать в принятии решений. В последнее десятилетие мы смогли достичь поставленных целей, но не так, как это представляли себе первопроходцы отрасли.

Неужели мы нашли способ имитировать человеческое мышление? До этого еще далеко. Наоборот, мы в корне пересмотрели основополагающее видение проблемы. Искусственный интеллект окружает нас повсюду, а его результативность сводится к big data и статистике: системы выполняют сложные вычисления на основе огромного количества данных. Мы смогли создать разум, но не такой, как наш. Мы все больше и больше полагаемся на эту новую форму интеллекта, и нам, возможно, придется пересмотреть и свое собственное мышление.

Не такие, как мы

Рик Рашид нервничал. И это было вполне объяснимо – выходя на сцену в 2012 году со своим обращением к нескольким тысячам ученых и студентов в Тяньцзине (Китай), он рисковал попасть в постыдную ситуацию. Рашид не говорил на китайском, а учитывая неудачное выступление его переводчика в прошлом, в этот раз мог произойти конфуз.

– Мы надеемся, что через несколько лет сможем преодолеть языковой барьер между людьми, – заявил основатель *Microsoft Research* своим слушателям. Возникла напряженная двухсекундная пауза, а затем из колонок послышался голос переводчика.

– Лично я уверен, что отсутствие языковых барьеров приведет нас к созданию лучшего мира, – продолжал Рашид. Еще одна пауза, и вновь прозвучал перевод на китайский. Рашид улыбнулся. Толпа аплодировала каждой реплике. Некоторые люди даже заплакали.

Столь восторженная реакция была вполне объяснимой: переводчик Рашида отлично показал себя. Каждое предложение было переведено идеально и понятно публике. Но самым впечатляющим было то, что речь Рашида переводил не человек.

Несмотря на все усилия ученых, когда-то выполнение по-

добной задачи выходило за пределы возможностей даже самого сложного искусственного интеллекта. На Дартмутской конференции в 1956 году и всех последующих мероприятиях были четко обозначены основные цели развития отрасли: машинный перевод, машинное зрение, понимание текста, распознавание речи, управление роботами и машинное обучение. Так появился целый список того, что мы хотели получить от ИИ.

На протяжении трех последующих десятилетий к исследованиям подключались мощнейшие ресурсы, однако ни один из пунктов списка не был достигнут. И лишь в конце 1990-х годов начали реализовываться многие прогнозы, предсказанные за 40 лет до этого. Но до новой волны успеха ИИ пришлось усвоить один очень важный и поучительный урок.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.