

NEW SCIENTIST
Лучшее от экспертов журнала



Machines that Think

*Everything You Need to Know about
the Coming Age of Artificial Intelligence*

NEW SCIENTIST

Искусственный интеллект

*Что стоит знать
о наступающей эпохе разумных машин*



ОГИ2

Издательство АСТ
Москва

УДК 004.8
ББК 32.813
И86

Machines that Think
Everything You Need to Know
about the Coming Age of Artificial Intelligence
First published in the English language by Hodder & Stoughton Limited.
Печатается с разрешения издательства Hodder & Stoughton Limited.

Нарушение прав автора, правообладателя, лицензиара влечет привлечение виновных к уголовной, административной и гражданской ответственности

Искусственный интеллект. Что стоит знать о наступающей эпохе разумных машин / под ред. Д. Хэвен ; пер. с англ. О. Д. Сайфудиновой. — Москва : Издательство АСТ, 2019. — 352 с. — (New Scientist. Лучшее от экспертов журнала).

ISBN 978-5-17-115608-4

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.

УДК 004.8
ББК 32.813

ISBN 978-5-17-115608-4
ISBN 978-1-4736-2965-3 (англ.)

© 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 внедре-

ния 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 Кида, Пол Маркс, Джастин Маллинс, Шон О'Нил, Сэнди Онг, Саймон Паркин, Самит Пол-Чоудхури, Тимоти Ревелл, Мэтт Рейнольдс, Дэвид Робсон, Авива 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 надеть «искрой человечности» машины.

Сравнение искусственного интеллекта с человеческим вызывает как радость, так и беспокойство. Насколько похожим 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. Собрав воедино мысли ведущих ис-

следователей 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 А являются Б, все А являются П, поэтому некоторые П являются Б* — может применяться к любым А, Б и П. Эта схема позволяет сделать правильное умозаключение вне зависимости от исходных данных. На основании формулы Аристоте-

ля 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 Элли, реагировала на окружающую обстановку снижением чувствительности к свету по мере разрядки аккумулятора. Столь сложное поведение делало ее непредсказуемой, на основании чего Уолтер проводил аналогию 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», опубликованной 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 ссыался 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 удалось 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.