Архитекторы интеллекта

Вся правда об искусственном интеллекте от его создателей

Иошуа Бенджио
Стюарт Рассел
Джеффри Хинтон
Ник Бостром
Ян Лекун
Фей-Фей Ли
Демис Хассабис
Эндрю Ын
Рана эль Калиуби
Рэймонд Курцвейл
Даниэла Рус
Джеймс Маника
Гари Маркус
Барбара Грош
Джуда Перл
Джефф Дин
Дафна Коллер
Дэвид Ферруччи
Родни Брукс
Синтия Бризил
Джошуа Тененбаум
Орен Этциони
Брайан Джонсон

Автор бестселлеров New York Times
Мартин Форд

ARCHITECTS OF INTELLIGENCE

THE TRUTH ABOUT AI FROM THE PEOPLE BUILDING IT

MARTIN FORD

Packt›
www.packt.com

Архитекторы интеллекта

ВСЯ ПРАВДА ОБ ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ ОТ ЕГО СОЗДАТЕЛЕЙ

Мартин Форд

ПИТЕР®

Санкт-Петербург • Москва • Екатеринбург • Воронеж
Нижний Новгород • Ростов-на-Дону
Самара • Минск

2020

ББК 32.813
УДК 004.8
Ф79

Форд Мартин

Ф79 Архитекторы интеллекта: вся правда об искусственном интеллекте от его создателей. — СПб.: Питер, 2020. — 416 с.: ил. — (Серия «Библиотека программиста»).

ISBN 978-5-4461-1254-8

Искусственный интеллект (ИИ) быстро переходит из области научной фантастики в повседневную жизнь. Современные устройства распознают человеческую речь, способны отвечать на вопросы и выполнять машинный перевод. В самых разных областях, от управления беспилотным автомобилем до диагностирования рака, применяются алгоритмы распознавания объектов на базе ИИ, возможности которых превосходят человеческие. Крупные медиакомпании используют роботизированную журналистику, создающую из собранных данных статьи, подобные авторским. Очевидно, что ИИ готов стать по-настоящему универсальной технологией, такой как электричество.

Какие подходы и технологии считаются наиболее перспективными? Какие крупные открытия возможны в ближайшие годы? Можно ли создать по-настоящему мыслящую машину или ИИ, сравнимый с человеческим, и как скоро? Какие риски и угрозы связаны с ИИ и как их избежать? Вызовет ли ИИ хаос в экономике и на рынке труда? Смогут ли суперинтеллектуальные машины выйти из-под контроля человека и превратиться в реальную угрозу?

Разумеется, предсказать будущее невозможно. Тем не менее эксперты знают о текущем состоянии технологий, а также об инновациях ближайшего будущего больше, чем кто бы то ни было. Вас ждут блестящие встречи с такими признанными авторитетами, как Р. Курцвейл, Д. Хассабис, Дж. Хинтон, Р. Брукс и многими другими.

16+ (В соответствии с Федеральным законом от 29 декабря 2010 г. № 436-ФЗ.)

ББК 32.813
УДК 004.8

ISBN 978-1789131512 англ.

ISBN 978-5-4461-1254-8

Оглавление

Посвящается Сяо-Сяо,
Элейн, Колину
и Тристану

Вступление

Мартин Форд

ФУТУРОЛОГ, ЭКСПЕРТ В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, КОНСУЛЬТАНТ ПО РАСЧЕТАМ ИНДЕКСА РОБОТОТЕХНИКИ RISE OF THE ROBOTS В SOCIETE GENERALE, ПРЕДПРИНИМАТЕЛЬ ИЗ КРЕМНИЕВОЙ ДОЛИНЫ, ЧЛЕН СОВЕТА ДИРЕКТОРОВ И ИНВЕСТОРОВ КОМПАНИИ GENESIS SYSTEMS

Мартин Форд — автор книг «Роботы наступают: развитие технологий и будущее без работы»[1] *(отмечена премией Financial Times & McKinsey Business Book of the Year, переведена более чем на 20 языков) и «Технологии, которые изменят мир»*[2]*. Писал о технологиях будущего для The New York Times, Fortune, Forbes, The Atlantic, The Washington Post, Harvard Business Review, The Guardian и The Financial Times. Выступал на многочисленных радио- и телешоу, в том числе на NPR, CNBC, CNN, MSNBC и PBS. Часто делает доклады о влиянии ИИ на экономику, рынок труда и общество будущего (наиболее известно выступление на конференции TED 2017 г.). Получил степень бакалавра в области вычислительной техники в Мичиганском университете в Анн-Арборе и степень MBA в Калифорнийском университете в Лос-Анджелесе (UCLA). В компании Genesis Systems участвует в разработке автоматизированных систем с автономным питанием, генерирующих воду непосредственно из воздуха.*

---

[1] *Форд М.* Роботы наступают: развитие технологий и будущее без работы / Пер. с англ. С. Чернина. — М.: Альпина нон-фикшн, 2016. — 429 с. : ил. — (Серия «Искусственный интеллект»).

[2] *Форд М.* Технологии, которые изменят мир / Пер. с англ. А. Кардаш. — М.: Манн, Иванов и Фербер, 2014. — 268 с.

Искусственный интеллект (ИИ) быстро переходит из области научной фантастики в повседневную жизнь. Современные устройства понимают человеческую речь, способны отвечать на вопросы и выполнять машинный перевод. В самых разных областях, от управления беспилотным автомобилем до диагностирования рака, применяются алгоритмы распознавания объектов на базе ИИ, возможности которых превосходят человеческие. Крупные медиакомпании используют роботизированную журналистику, создающую из собранных данных статьи, подобные авторским. Очевидно, что ИИ готов стать одним из важнейших факторов, формирующих наш мир, являясь по-настоящему универсальной технологией, такой как электричество.

В последние годы в СМИ широко освещаются достижения в области ИИ. Бесчисленные статьи, книги, документальные фильмы и телепрограммы предсказывают новую эру, мешая в одну кучу анализ фактических данных с ажиотажем, спекуляциями и даже нагнетанием паники. Говорят, что через несколько лет дороги полностью захватят беспилотные автомобили, оставив без работы водителей грузовиков и такси. В некоторых алгоритмах машинного обучения обнаружили признаки дискриминации по расовому и половому признакам. Неясно, как повлияет на конфиденциальность распознавание лиц. Роботы могут стать оружием. А обладающие интеллектом машины представляют угрозу существованию человечества. Свое мнение озвучивают многие общественные деятели, которые не являются экспертами в сфере ИИ. Радикальнее всего выступил Илон Маск, заявивший, что разработки ИИ сродни призыву демонов и опаснее ядерного оружия. Даже Генри Киссинджер и Стивен Хокинг публиковали мрачные прогнозы.

Поэтому мне хотелось бы рассказать о том, что такое ИИ, и осветить связанные с ним возможности и риски. Для этого я провел серию интервью с выдающимися учеными и предпринимателями, занимающимися ИИ. Многие из них лично повлияли на трансформацию окружающего нас мира; другие — основали компании,

которые расширяют границы ИИ, робототехники и машинного обучения.

Разумеется, сформированный мной список субъективен — в развитии ИИ участвует множество профессионалов. Но я уверен, что почти любой человек, обладающий глубокими знаниями в этой области, поддержит мой выбор. Всех этих людей можно без преувеличения назвать творцами ИИ, приближающими начало новой научно-технической революции.

В интервью я старался задать наиболее острые вопросы, появившиеся в процессе развития ИИ. Какие подходы и технологии считаются наиболее перспективными? Какие крупные открытия возможны в ближайшие годы? Можно ли создать по-настоящему мыслящую машину или ИИ, сравнимый с человеческим, и как скоро? Какие риски и угрозы связаны с ИИ и как их избежать? Требуется ли для этой области государственное регулирование? Вызовет ли ИИ хаос в экономике и на рынке труда? Смогут ли суперинтеллектуальные машины выйти из-под контроля человека и превратиться в реальную угрозу? Нужно ли беспокоиться о «гонке вооружений» в области ИИ?

Разумеется, предсказать будущее невозможно. Тем не менее эксперты знают о текущем состоянии технологий, а также об инновациях ближайшего будущего больше, чем кто бы то ни было. Поэтому их мысли и мнения заслуживают внимания. Помимо ИИ, мы обсудили образование, карьеру и исследовательские интересы, поэтому чтение будет увлекательным и вдохновляющим.

Искусственный интеллект — это широкая область исследований, сопряженная с множеством дополнительных дисциплин. Многие из моих собеседников совмещали работу в нескольких областях. Сейчас я кратко расскажу, как опрошенные относятся к наиболее важным инновациям в исследованиях ИИ и задачам будущего. Основная информация о каждом из них будет приведена в начале соответствующего интервью.

Подавляющее большинство достижений сферы ИИ последнего десятилетия — от распознавания лиц до машинного перевода и победы в игре го — основаны на технологии глубокого обучения, или глубоких нейронных сетей. Искусственные нейронные сети, в которых программно эмулируется структура и взаимодействие нейронов головного мозга, появились примерно в 1950-х гг. Простые версии этих сетей могли решать элементарные задачи по распознаванию объектов на изображениях, что сначала вызывало сильный энтузиазм. Однако к 1960 г., частично из-за критики Марвина Минского — одного из пионеров ИИ, — нейронные сети потеряли популярность, а им на смену пришли другие подходы.

В течение примерно 20 лет, начиная с 1980-х гг., небольшая группа исследователей продолжала верить в технологию нейронных сетей и продвигать ее. Среди них выделялись Джеффри Хинтон (Geoffrey Hinton), Иошуа Бенджио (Yoshua Bengio) и Ян Лекун (Yann LeCun). Они не только внесли вклад в лежащую в основе глубокого обучения математическую теорию, но и первыми стали продвигать технологию «глубоких» сетей с несколькими слоями искусственных нейронов. Им удалось донести идею нейронных сетей до времен экспоненциального роста вычислительных мощностей и увеличения объема доступных данных. В 2012 г. команда аспирантов Хинтона из Университета Торонто победила в конкурсе по распознаванию объектов на изображениях.

После этого события глубокое обучение стало общедоступным. Большинство крупных технологических компаний — Google, Facebook, Microsoft, Amazon, Apple, Baidu и Tencent — инвестировали огромные суммы в новую технологию, чтобы использовать ее в своем бизнесе. Разработчики микропроцессорных и графических чипов (GPU), такие как NVIDIA и Intel, переорганизовали бизнес под создание оборудования, оптимизированного для нейронных сетей. Именно глубокое обучение сегодня раскрывает сферу ИИ.

Такие ученые, как Эндрю Ын (Andrew Ng), Фей-Фей Ли (Fei-Fei Li), Джефф Дин (Jeff Dean) и Демис Хассабис (Demis Hassabis), используют современные нейронные сети в таких областях, как поисковые системы, компьютерное зрение, беспилотные автомобили и универсальный ИИ. Это признанные лидеры в области преподавания, управления и предпринимательства на базе технологии нейронных сетей.

Однако глубокое обучение подвергается критике. Ряд ученых считают его «одним из инструментов в наборе», утверждая, что для дальнейшего прогресса нужны идеи из других областей. Барбара Грош (Barbara Grosz) и Дэвид Ферруччи (David Ferrucci) занимаются проблемами понимания естественного языка. Гари Маркус (Gary Marcus) и Джош Тененбаум (Josh Tenenbaum) изучают человеческое познание. Орен Этциони (Oren Etzioni), Стюарт Рассел (Stuart Russell) и Дафна Коллер (Daphne Koller) специализируются на вероятностных методах. Джуда Перл (Judea Pearl) за работу по вероятностным (или байесовским) подходам к ИИ и машинному обучению получил премию Тьюринга.

Сфера робототехники также развивается благодаря таким ученым, как Родни Брукс (Rodney Brooks), Даниэла Рус (Daniela Rus) и Синтия Бризил (Cynthia Breazeal). Бризил вместе с Раной эль Калиуби (Rana El-Kaliouby) — первопроходцы в построении систем, умеющих распознавать эмоции, реагировать на них и вступать в социальные взаимодействия с людьми. Брайан Джонсон (Bryan Johnson) основал компанию *Kernel*, направляющую технологии ИИ в развитие человека.

По моему мнению, особый интерес представляют три основные темы, поэтому они будут рассматриваться в каждом интервью. Первая касается автоматизации человеческого труда, ведущей к росту безработицы. Глубже всего эту тему раскрыл Джеймс Маника (James Manyika) — глава Глобального института McKinsey (MGI), где активно исследуется влияние технологий на рынок труда.

Второй вопрос, который я задавал всем, касается ИИ, сравнимого с человеческим. Это так называемый сильный ИИ (artificial general intelligence, AGI), который был недостижимой мечтой. Демис Хассабис рассказал, что предпринимается в компании DeepMind, которая является крупнейшей и наиболее финансируемой инициативой по исследованиям сильного ИИ. Дэвид Ферруччи, руководивший разработкой суперкомпьютера IBM Watson, — генеральный директор стартапа Elemental Cognition, — описал создание сильного ИИ путем эффективного применения понимания языка. Важные идеи высказал и Рэймонд Курцвейл (Raymond Kurzweil) — автор книги *Singularity is Near* («Сингулярность уже близка»), в настоящее время руководящий проектом Google, связанным с обработкой естественного языка.

Всем я задал вопрос: «К какому году с вероятностью 50 % будет создан ИИ уровня человеческого?» Большинство предпочло поделиться своими предположениями анонимно. Двое из опрошенных выразили желание официально поделиться своей точкой зрения. Результаты этого опроса приведены в конце книги. Вы увидите, как мнения по важным темам зачастую кардинально расходятся, что представляет собой один из наиболее интересных аспектов данной книги.

Третья обсуждаемая тема связана с последствиями прогресса в области ИИ, ожидаемыми как в ближайшем, так и в отдаленном будущем. Становится очевидной проблема уязвимости взаимосвязанных автономных систем к атакам через интернет. Также выявлена предрасположенность алгоритмов машинного обучения к предвзятости по расовому и половому признакам. Многие из моих собеседников подчеркнули важность решения этой проблемы и рассказали об исследованиях в этой области. Некоторые дали оптимистический прогноз, предположив, что ИИ поможет нам избавиться от предвзятости и дискриминации.

Многих волнует опасность появления полностью автономного оружия. В сообществе исследователей ИИ существует мнение,

что роботы или дроны, способные убивать без контроля человека, в конечном итоге могут стать не менее опасными, чем биологическое или химическое оружие. В июле 2018 г. более 160 компаний и 2400 исследователей (среди которых есть мои собеседники) подписали соглашение о запрете производства смертоносных алгоритмов[1].

Более отдаленной и умозрительной является проблема несоответствия собственных целей сильного ИИ с желаниями человека. Этой темы касались почти все мои собеседники. Чтобы адекватно и рационально осветить эту проблему, я поговорил с Ником Бостромом (Nick Bostrom) из оксфордского Института будущего человечества (Future of humanity institute, FHI) — автором бестселлера «Искусственный интеллект: этапы, угрозы, стратегии»[2], в котором тщательно рассматриваются потенциальные риски, связанные с машинами, интеллектуально превосходящими человека.

Создатели интеллекта

Интервью для этой книги проводились с февраля по август 2018 г. Практически все они длились не меньше часа, а некоторые существенно дольше. Записанные, транскрибированные, а затем отредактированные командой издательства Packt тексты я дал своим собеседникам на проверку. Уверен, что книга верно отражает мысли респондентов.

Эксперты, с которыми я общался, имеют разное происхождение и сотрудничают с разными компаниями. Но вы быстро обнаружите, насколько сильно влияние Google на сообщества, связанные с ИИ. Из двадцати трех специалистов у семи есть или были связи с Google или холдингом Alphabet. Много талантливых людей работает в Массачусетском технологическом институте (MIT)

[1] https://futureoflife.org/lethal-autonomous-weapons-pledge/

[2] *Бостром Н.* Искусственный интеллект: этапы, угрозы, стратегии / Пер. с англ. С. Филина. — М.: Манн, Иванов и Фербер, 2016. — 490 с. : ил.

и Стэнфорде. Джеффри Хинтон и Иошуа Бенджио представляют университеты Торонто и Монреаля соответственно, а правительство Канады ведет четкую промышленную политику, ориентированную на робототехнику и ИИ. В Соединенных Штатах работали 19 из 23 опрошенных, но больше половины из них родились за пределами США: в Австралии, Китае, Египте, Франции, Израиле, Родезии (ныне Зимбабве), Румынии и Великобритании. Это ярко иллюстрирует роль иммиграции квалифицированных кадров в технологическом лидерстве США.

Проводя интервью, я все время помнил, что книгу будут читать самые разные люди, от специалистов по теории вычислительных машин и систем до менеджеров и инвесторов. Но самая важная часть аудитории — молодые люди, которые могут задуматься о карьере в области ИИ. Сейчас в ней наблюдается дефицит кадров, особенно специалистов с навыками глубокого обучения, что дает возможность для хорошего карьерного роста. В настоящее время прилагаются усилия по привлечению в отрасль талантливых специалистов, и уже широко признается необходимость профессиональной интеграции.

Около четверти опрошенных мной — женщины. Здесь их доля выше, чем в сфере ИИ и машинного обучения в целом. Согласно недавним исследованиям, женщины составляют примерно 12 % ведущих сотрудников в области машинного обучения[1]. В процессе интервью многие подчеркивали необходимость увеличения доли как женщин, так и представителей меньшинств.

Одна из моих собеседниц уделяет особое внимание многообразию в области ИИ. Фей-Фей Ли из Стэнфорда — соучредитель AI4ALL[2], устраивающей летние учебные лагеря для старшеклассников из мало представленных в этой сфере групп. AI4ALL получила поддержку отрасли, а также грант от Google, и теперь такие про-

[1] https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance

[2] http://ai-4-all.org/

граммы проводятся в шести американских университетах. В этом направлении еще многое предстоит сделать, но основания для оптимистических прогнозов уже есть.

Хотя книга рассчитана на широкий круг читателей, в тексте будут встречаться специальные понятия и термины. Если вы ранее ничего не знали об ИИ, то я рад, что вас познакомят с ним ведущие специалисты, и рекомендую вам начать с краткого словаря, приведенного ниже. В интервью Стюарта Рассела — соавтора ведущего учебника по ИИ — вы найдете объяснение ключевых концепций области.

Возможность взять эти интервью была для меня честью. Надеюсь, вы тоже увидите в моих собеседниках вдумчивость, умение рассказывать и глубокую приверженность идее работы на благо человечества. Чего в книге нет, так это единодушия. Она наполнена разнообразными, зачастую резко противоречивыми представлениями, мнениями и прогнозами. Понятно только одно: ИИ — широко открытое пространство. Можно строить предположения о природе будущих инноваций, скорости их появления и конкретных вариантах их применения. Именно из-за этой комбинации потенциальной разрушительности с фундаментальной неопределенностью необходим содержательный и всеобъемлющий разговор о будущем ИИ и его влиянии на наш образ жизни. Надеюсь, моя книга внесет в него свой вклад.

Краткий словарь терминов

В нескольких интервью углубленно рассматриваются методы, используемые в сфере ИИ. Для понимания материала специальных знаний не требуется, но встречающиеся термины желательно знать. Вот объяснение наиболее важных из них. Если вы сочтете какой-то раздел технически сложным и запутанным, просто пропустите его и переходите к следующему.

Машинное обучение (**machine learning**) — раздел ИИ о методах построения алгоритмов, способных обучаться на данных. Другими словами, алгоритмы машинного обучения — это компьютерные программы, которые, по сути, программируют сами себя, просматривая информацию. Раньше считалось, что «компьютеры совершают только те действия, которые были запрограммированы», но эта ситуация меняется. Среди многочисленных типов алгоритмов машинного обучения самый революционный (и привлекающий всеобщее внимание) — это глубокое обучение.

Глубокое обучение (**deep learning**) — вид машинного обучения, в котором используются глубокие (или многоуровневые) **искусственные нейронные сети** (**artificial neural networks**), то есть программное обеспечение, имитирующее работу нейронов мозга. Глубокое обучение послужило основной движущей силой развития ИИ.

Есть и другие термины, которые, скорее всего, новичкам покажутся сложными. Без их глубокого понимания вполне можно обойтись, но краткое пояснение лишним не будет.

Метод обратного распространения ошибки (**backpropagation**) — алгоритм, используемый в системах глубокого обучения. Информация, поступающая в нейронную сеть, распространяется обратно через слои нейронов, вызывая у некоторых из них изменение настроек — весов (см. ниже «Обучение с учителем»). Так постепенно сеть находит правильный ответ. В 1986 г. Джеффри Хинтон стал соавтором первой полноценной статьи на эту тему, о чем более подробно вы узнаете из интервью с ним.

Еще более непонятный термин — **градиентный спуск** (**gradient descent**) — относится к математической технике, которую алгоритм обратного распространения использует для уменьшения ошибки в процессе обучения сети.

Встречаются в книге и термины, относящиеся к типам или конфигурациям нейронных сетей: **рекуррентные** (**recurrent**)

и **сверточные** (**convolutional**) **сети**, а также **машины Больцмана** (**Boltzmann machines**). Различия обычно сводятся к способам связи нейронов. Детальное рассмотрение этих понятий выходит за рамки книги. Тем не менее я попросил объяснить их Яна Лекуна — изобретателя сверточной архитектуры, которая широко используется в распознавании объектов на изображениях.

Термин **байесовский** (**bayesian**) можно перевести как «вероятностный». Он встречается в таких сочетаниях, как «байесовские методы машинного обучения» или «байесовские сети». Они относятся к алгоритмам, которые используют вероятностные зависимости. Термин назван в честь священника Томаса Байеса (1701–1761), который сформулировал способ обновления вероятности события после возникновения другого, статистически взаимозависимого с ним. Байесовские методы очень популярны как среди специалистов по теории вычислительных машин и систем, так и среди ученых, моделирующих человеческое познание. Больше всего по этой теме рассказал Джуда Перл.

Способы обучения ИИ-систем

Существуют разные типы машинного обучения. Решающую роль в развитии искусственного интеллекта играют инновации, то есть новые способы обучения систем ИИ.

При **обучении с учителем** (**supervised learning**) алгоритму передаются структурированные, классифицированные и снабженные метками данные. Например, чтобы научить систему глубокого обучения распознавать на снимках собак, ей нужно предоставить много тысяч (или даже миллионов) изображений этого животного с меткой «собака». Кроме того, потребуется огромное количество изображений без собаки с меткой «нет собаки». После обучения можно показывать системе новые фотографии, и она будет определять наличие на них собаки на уровне, превосходящем возможности обычного человека.

Обучение с учителем — наиболее распространенный метод, применяемый в современных системах ИИ. На его долю приходится около 95 % практических приложений. Именно оно послужило основой машинного перевода (после обучения на миллионах предварительно переведенных документов) и ИИ-систем диагностики (после обучения на снимках с пометками «рак» и «нет рака»). К сожалению, для такого обучения требуются огромные объемы маркированных данных. Именно поэтому лидирующее положение в технологии глубокого обучения занимают такие компании, как Google, Amazon и Facebook.

Обучение с подкреплением (**reinforcement learning**), по сути, представляет собой обучение на практике или методом проб и ошибок. Система учится не на правильных размеченных данных, а самостоятельно ищет решение, получая подкрепление в случае успеха. Это напоминает дрессировку животных, которым в случае правильных действий дается кусочек вкусной еды. Именно обучение с подкреплением применялось для построения систем ИИ, играющих в игры. Из интервью с Демисом Хассабисом вы узнаете, что компания DeepMind использовала этот тип обучения для разработки компьютерной системы AlphaGo.

Проблема обучения по этому алгоритму заключается в необходимости огромного количества тренировочных попыток. Поэтому он применяется в основном для игр или для задач, которые можно воспроизводить на компьютере с высокой скоростью. Обучение с подкреплением можно использовать при разработке беспилотных автомобилей, но не для их эксплуатации на реальных дорогах. Виртуальные машины обучаются в искусственной среде, а после завершения обучения программное обеспечение устанавливается на реальные автомобили.

Обучение без учителя (**unsupervised learning**) обеспечивает непосредственное обучение на поступающих из окружающей среды неструктурированных данных. Именно так учатся люди. Например, дети учатся говорить, слушая речь родителей. Раз-

умеется, человек использует и другие типы обучения, но самым характерным для него остается наблюдение и неконтролируемое взаимодействие с окружающей средой.

Обучение без учителя — один из наиболее многообещающих путей развития ИИ. Только представьте системы, умеющие обучаться сами без подготовки данных. Но их разработка — одна из самых сложных задач. Ее решение станет важной точкой на пути к созданию сильного ИИ.

Термин **сильный ИИ** обозначает истинно мыслящую машину, изначальную цель создания ИИ. Еще его называют интеллектом, сравнимым с человеческим разумом. Примеры сильного ИИ можно наблюдать в научной фантастике: компьютер HAL 9000 из «Космической одиссеи», главный компьютер космического корабля «Энтерпрайз» (или Дэйта) из «Звездного пути», андроид C3PO из «Звездных войн» и агент Смит из «Матрицы». Все эти вымышленные системы могли пройти **тест Тьюринга** (**Turing test**), то есть вести беседу как человек. Этот тест был предложен Аланом Тьюрингом в статье 1950 г. «Вычислительные машины и разум»[1], которую можно считать основополагающей работой в области ИИ.

Есть вероятность, что когда-нибудь появится **суперинтеллект** (**superintelligence**), или машина, превосходящая интеллектуальные способности любого человека. Это может произойти в результате простого увеличения аппаратных мощностей и быть ускорено самосовершенствованием этой машины. Так она запустит «рекурсивный цикл улучшения» или «быстрый интеллектуальный взлет», создавая проблему «выравнивания», если вступит в противоречие с интересам человека.

[1] Тьюринг А. Вычислительные машины и разум / Пер. с англ. К. Королева. — М.: АСТ, 2018. — 128 с. — (Серия «Эксклюзивная классика»).

> *ИИ, который существует сейчас и может появиться в обозримом будущем, не понимает и не чувствует нормы морали.*

Иошуа Бенджио

ДИРЕКТОР МОНРЕАЛЬСКОГО ИНСТИТУТА АЛГОРИТМОВ ОБУЧЕНИЯ (MILA), ДОКТОР COMPUTER SCIENCE, ПРОФЕССОР КАФЕДРЫ ИНФОРМАТИКИ И МАТЕМАТИЧЕСКИХ МЕТОДОВ МОНРЕАЛЬСКОГО УНИВЕРСИТЕТА, СОРУКОВОДИТЕЛЬ ПРОЕКТА LEARNING IN MACHINES & BRAINS КАНАДСКОГО ИНСТИТУТА ПЕРСПЕКТИВНЫХ ИССЛЕДОВАНИЙ (CIFAR)

Иошуа Бенджио широко известен как один из пионеров глубокого обучения. Он активно продвигал исследования нейронных сетей, в частности обучение без учителя, и стал соавтором книги «Глубокое обучение»[1] *— одним из основных учебников по одноименному предмету.*

[1] *Бенджио И., Гудфеллоу Я., Курвилль А.* Глубокое обучение / Пер. с англ. А. Слинкина. — М.: ДМК пресс, 2017. — 652 с.: ил. Книга бесплатно доступна по адресу https://www.deeplearningbook.org.

Мартин Форд: Вы играете ведущую роль в исследованиях ИИ, поэтому начать мне хотелось бы с вопроса о том, какие исследовательские проблемы стоят на пути к сильному ИИ.

Иошуа Бенджио: До создания ИИ, сравнимого с человеческим, нам еще очень далеко. Нужно понять, к примеру, почему невозможно создать машину, которая понимала бы окружающую действительность так же, как человек. Чего нам не хватает: обучающих данных или вычислительных мощностей? Многие считают, что причина состоит в отсутствии необходимых базовых компонентов, например, умения видеть причинно-следственные связи в данных, которое позволяет делать обобщения и находить правильные ответы в условиях, отличных от тренировочных.

Человек может представить, как он переживет новый для себя опыт. Например, если вы никогда не попадали в автомобильную аварию, вы все равно сможете прокрутить у себя в голове такую ситуацию и принять правильное решение. Обучение с учителем помогает компьютеру находить статистические закономерности в поставляемых данных, которые заранее классифицированы и размечены людьми.

Многие исследования пока не дали значимых результатов. Компьютер не может автономно приобретать знания о мире, воздействовать на него и наблюдать результат воздействия. Ответы на вопрос, как это реализовать, ищем не только мы.

М. Ф.: Какие проекты в настоящее время можно считать первостепенными в области глубокого обучения? Мне первым делом вспоминается программа AlphaZero. Есть ли другие?

И. Б.: На мой взгляд, из множества интересных проектов наиболее перспективны те, в которых агент в виртуальном мире пытается решать задачи, попутно изучая все с ними связанное. Такими проектами занимаемся мы в MILA, а также компании DeepMind, OpenAI, Университет Беркли, Facebook и Google в рамках проекта Google Brain. Это новые горизонты.

Но это долговременные исследования. Мы работаем не над конкретными вариантами применения глубокого обучения, а над тем, как научить агента осмысливать окружающую среду, говорить и понимать так называемый обоснованный язык (grounded language).

М. Ф.: Что означает этот термин?

И. Б.: Раньше компьютеры обучались языку, знакомясь с множеством текстов. Причем они достигали понимания только через связь слова с называемой им реалией. В отличие от робота, человек может сопоставить слово не только с объектом из реального мира, но и с вариантами изображения этого объекта.

Многочисленные исследования в области обучения обоснованному языку сводятся к попыткам научить роботов понимать язык хотя бы на уровне отдельных слов и выражений и реагировать соответствующим образом. Это очень интересное направление, необходимое для реализации таких вещей, как диалог с роботами, личные помощники и т. п.

М. Ф.: То есть, по сути, идея состоит в том, чтобы дать агенту свободу в смоделированной среде, позволив ему учиться, как это делают дети?

И. Б.: Именно так. Более того, мы пользуемся результатами исследований в области детского развития и изучаем, какие этапы проходит новорожденный в первые месяцы жизни, постепенно приобретая представления о мире. До сих пор не совсем понятно, какие умения являются врожденными, а какие получены путем изучения.

Несколько лет назад я предложил для машинного обучения практику, которая используется при дрессировке животных — обучение по плану (curriculum learning). Обучающие примеры в этом случае демонстрируются не произвольно, а в последовательности, целесообразной для обучения. Процесс начинается с простых

концепций, которые после их освоения учеником можно использовать как «кирпичики» для объяснения более сложных понятий.

М. Ф.: Я бы хотел поговорить о работе над сильным ИИ. Очевидно, что важной составляющей этого процесса вы считаете обучение без учителя. Что еще необходимо сделать?

И. Б.: Мой друг Ян Лекун сравнивает этот процесс с подъемом на гору. Сначала все радуются, насколько высоко забрались, но по мере приближения к вершине встречается множество других гор. Сейчас при разработке сильного ИИ четко видна ограниченность используемых подходов. Пока мы искали способы обучения более глубоких сетей, взбираясь на первую гору, создаваемые системы исследовались очень узко — на том этапе было важно просто подняться на несколько шагов вверх.

Как только применяемые техники обучения дали первые удовлетворительные результаты — мы приблизились к вершине первой горы, — стали заметны ограничения. И это следующая гора, которую нужно будет покорять. Поэтому невозможно сказать, сколько еще открытий потребуется.

М. Ф.: А вы можете хотя бы примерно оценить количество гор? Или период времени, который потребуется на создание сильного ИИ? Просто поделитесь своими прогнозами.

И. Б.: Не вижу смысла говорить о сроках. Невозможно предсказать, когда именно будет открыта дверь, от которой у нас нет ключа. Могу только заверить, что в ближайшие годы никаких прорывов не будет.

М. Ф.: Считаете ли вы перспективными глубокое обучение и нейронные сети в целом?

И. Б.: Да, многолетний прогресс в области глубокого обучения и нейронных сетей означает, что открытые концепции будут активно использоваться и дальше. Возможно, именно они помо-

гут понять, каким образом мозг животных и человека осваивает сложные понятия. Но этого недостаточно для создания сильного ИИ. В настоящее время мы видим ограниченность имеющихся систем и собираемся улучшать и развивать их.

М. Ф.: Я знаю, что Институт искусственного интеллекта Пола Аллена (AI2) работает над проектом Mosaic, в рамках которого компьютеру пытаются помочь обрести разум. Считаете ли вы, что это важная задача? Ведь, возможно, разум рождается в процессе обучения?

И. Б.: Я уверен, что он возникает именно как результат обучения. Разум не может появиться только потому, что кто-то положил вам в голову какие-то знания. По крайней мере, у людей так.

М. Ф.: Глубокое обучение — основной путь к созданию сильного ИИ или потребуются гибридные системы?

И. Б.: Изначально ИИ был условным понятием, ни о каком обучении речи не шло. В центре внимания была способность машины делать последовательные выводы и объединять фрагменты информации. А глубокое обучение нейронных сетей можно назвать познанием снизу вверх. Все начинается с восприятия, в котором мы закрепляем понимание мира машиной. Затем можно строить распределенные представления и фиксировать связи между множеством переменных.

Отношения между такими переменными мы с братом изучали в 1999 г., что дало толчок к появлению в области естественного языка таких подходов, как векторное представление слов или распределенные представления слов и предложений. В них слово описывается характером активности в мозге или набором чисел. Слова со сходными значениями связываются со сходными числовыми комбинациями.

В настоящее время на базе этих подходов пытаются решать классические проблемы ИИ, связанные с умением рассуждать и по-

нимать, программировать и планировать. «Строительные блоки», обнаруженные при изучении восприятия, сейчас пробуют распространять на когнитивные задачи более высокого уровня (психологи называют это действиями Системы 2). Я полагаю, именно таким способом мы будем двигаться к сильному ИИ. Это нельзя назвать гибридной системой; скорее, мы пытаемся работать над классическим ИИ, используя как строительный материал концепции из глубокого обучения. Можно сказать, что требуются альтернативные пути достижения цели.

М. Ф.: То есть вы считаете, что все сведется к нейронным сетям с различными архитектурами?

И. Б.: Да. Ведь человеческий мозг состоит из нейронных сетей. Нужно придумать архитектуры и обучающие техники, позволяющие решать задачи, поставленные перед классическим ИИ.

М. Ф.: Обучения и тренировки будет достаточно или потребуется какая-то дополнительная структура?

И. Б.: Она уже существует, просто отличается от привычной структуры представления знаний, которую мы наблюдаем в энциклопедиях или формулах. Она имеет архитектуру нейронной сети и довольно широкие допущения по поводу окружающего мира и вершины собственных возможностей. Чтобы реализовывать в нейронной сети механизм внимания, такая структура требует большого количества предварительных знаний. Оказывается, данные имеют решающее значение для таких вещей, как машинный перевод.

Уже существует множество предположений в разных предметных областях о мире и о внедряемой функции, которые в виде архитектур и целей содержались в технологии глубокого обучения. Именно этому посвящено большинство современных научных работ.

М. Ф.: Говорят, что новорожденные развивают навык распознавания лиц с первых дней жизни. Очевидно, что это возможно

благодаря некой структуре в мозге. Это не просто реакция нейронов на пикселы.

И. Б.: Ошибаетесь! Это именно реакция нейронов на пикселы, кроме того, в мозге ребенка присутствует особая структура, которая распознает нечто круглое с двумя точками внутри.

М. Ф.: Я считаю, что она существует с момента рождения.

И. Б.: Разумеется. И все то, что мы проектируем в нейронных сетях, тоже существует с самого начала. Работа исследователя в области глубокого обучения напоминает процесс эволюции. Знания вводятся как в виде структуры, так и через обучение.

При желании можно создать нечто, позволяющее сети распознавать лица, но в этом нет смысла, так как ИИ быстро обучается. Поэтому мы работаем над решением более сложных проблем.

Никто не говорит об отсутствии врожденных знаний у людей, детей и животных. Более того, у большинства животных знания исключительно врожденные. Муравью не приходится долго учиться, он действует в соответствии с заложенной в него программой. Но чем выше существо в иерархии интеллекта, тем большую роль в его жизнедеятельности начинает играть обучение. Человека отличает именно соотношение врожденных и приобретенных навыков.

М. Ф.: Я бы хотел уточнить некоторые из этих концепций. В 1980-е гг., после периода забвения, снова появился интерес к нейронным сетям, но речи о множестве слоев и глубине еще не шло. Вы участвовали в развитии глубокого обучения. Не могли бы вы простыми словами объяснить, что это такое?

И. Б.: Глубокое обучение — это совокупность методов машинного обучения. Но если в случае классического машинного обучения компьютеры учатся по прецедентам, глубокое обучение больше напоминает процесс, происходящий в мозге человека.

Эти методы работы над ИИ появились как продолжение более раннего изучения нейронных сетей. Слово «глубокие» указывает на появление у сетей дополнительных уровней, со своими вариантами представления информации. Мы надеемся, что углубление сетей позволит машине представлять более абстрактные вещи.

М. Ф.: То есть под слоями вы подразумеваете уровни абстракции? И если в качестве примера взять изображение, то первым уровнем будут пикселы, затем контуры и т. д.?

И. Б.: Да, все правильно.

М. Ф.: Правда ли то, что компьютеры до сих пор не понимают, что такое объект?

И. Б.: До некоторой степени компьютер понимает. Скажем, кошка понимает, что такое дверь, но не так, как человек. Даже люди обладают разными уровнями понимания многих вещей, а наука призвана углубить это понимание. Люди интерпретируют образы в контексте трехмерного мира благодаря стереоскопическому зрению и опыту познания. Человек получает не визуальную, а физическую модель объекта. Компьютер интерпретирует изображения на примитивном уровне, но для множества приложений этого достаточно.

М. Ф.: Правда ли, что глубокое обучение стало возможным благодаря методу обратного распространения ошибки, основная идея которого состоит в том, что информацию об ошибке можно отправить от выходов сети к ее входам, корректируя каждый слой в зависимости от конечного результата?

И. Б.: Да, метод обратного распространения стал краеугольным камнем успехов глубокого обучения. Он позволяет присваивать данным коэффициенты доверия (credit assignment), то есть рассчитывать, как для корректного поведения всей сети должны измениться внутренние нейроны. В контексте нейронных сетей

об этом методе заговорили в начале 1980-х гг., когда я только начинал работать самостоятельно. Одновременно с Яном Лекуном метод развивали Джеффри Хинтон и Дэвид Румельхарт (David Rumelhart). Идея не новая, но примерно до 2006 г. особых успехов в обучении глубоких сетей не наблюдалось. Сейчас мы имеем механизм внимания, память и способность не только классифицировать, но и генерировать изображения.

М. Ф.: Существуют ли аналоги обратного распространения в человеческом мозге?

И. Б.: Хороший вопрос. Дело в том, что нейронные сети не пытаются скопировать мозг, хотя и появились как попытка смоделировать некоторые происходящие в нем процессы. Мы полностью не понимаем, как работает мозг. Нейробиологи пока не соединили результаты своих наблюдений в общую картину. Возможно, наша работа сможет дать доступную для проверки гипотезу. Ведь метод обратного распространения до сих пор считался уделом компьютеров, но не человеческого мозга. Прекрасные результаты, которые он дает, заставляют подозревать, что, возможно, мозг умеет проделывать похожие штуки. Я участвую в исследованиях, которые могут дать ответ на этот вопрос.

М. Ф.: В период «зимы ИИ», когда общий интерес к нему угас, вы вместе с Джеффри Хинтоном и Яном Лекуном продолжали свои исследования. Как вам удалось добиться таких успехов, как сейчас?

И. Б.: К концу 1990-х гг. нейронные сети вышли из моды, и ими практически никто не занимался. Но моя интуиция говорила, что мы упускаем что-то важное. Ведь благодаря композиционной структуре они могли представить богатую информацию о данных, базируясь на множестве «строительных блоков» — нейронов и их слоев. Лично меня это привело к лингвистическим моделям, то есть к нейронным сетям, которые моделировали текст, используя векторные представления слов. Каждое слово в них связано с набором чисел, соответствующих различным атрибутам,

которые изучаются машиной автономно. Тогда этот подход не получил широкого распространения, но в настоящее время эти идеи используются почти во всем, что связано с моделированием языка на основе данных.

Обучать глубокие сети мы не умели, но проблему решил Джеффри Хинтон своей работой по быстрым алгоритмам обучения ограниченной машины Больцмана (restricted Boltzmann machine, RBM). В моей лаборатории велась работа над связанными с ней автокодировщиками, которые дали начало таким моделям, как генеративно-состязательные сети (generative adversarial networks). Благодаря им появилась возможность обучения более глубоких сетей.

М. Ф.: А что такое автокодировщик?

И. Б.: Это специальная архитектура, состоящая из двух частей: кодировщика и декодера. То, что кодировщик сжал — декодер восстанавливал, причем так, чтобы выход был максимально близок к оригиналу. Автокодировщики превращали входную необработанную информацию в более абстрактное представление, в котором проще было выделить семантический аспект. Затем декодер восстанавливал по этой высокоуровневой абстракции исходные данные. Это были первые работы по глубокому обучению.

Через несколько лет мы обнаружили, что для обучения глубоких сетей достаточно изменения нелинейности. Вместе с одним из моих студентов, который работал с нейробиологами, мы решили попробовать блоки линейной ректификации (rectified linear unit, ReLU). Это пример копирования работы человеческого мозга.

М. Ф.: И к каким результатам это привело?

И. Б.: Раньше для активации нейронных сетей использовали сигмоиду, но оказалось, что с функцией ReLU гораздо проще обучать глубокие сети с большим количеством уровней. Переход случился

примерно в 2010 г. Появилась огромная база данных ImageNet, предназначенная для отработки и тестирования методов распознавания объектов на изображениях и машинного зрения. Чтобы заставить людей поверить в методы глубокого обучения, нужно было показать хорошие результаты на примере этой базы. Это смогла сделать группа Джеффри Хинтона, которая использовала в качестве основы работы Яна Лекуна, посвященные сверточным сетям. В 2012 г. эти новые архитектуры позволили значительно улучшить существующие методы. За пару лет на эти сети переключились все, кто занимался компьютерным зрением.

М. Ф.: То есть именно в этот момент началось настоящее глубокое обучение?

И. Б.: Нет, совокупность факторов, ускоривших глубокое обучение, целиком сложилась только к 2014 г.

М. Ф.: То есть к моменту, когда этим занялись не только университеты, но и такие компании, как Google, Facebook и Baidu?

И. Б.: Именно так. Процесс ускорения начался чуть раньше, примерно в 2010 г., благодаря таким компаниям, как Google, IBM и Microsoft, которые работали над нейронными сетями для распознавания речи. Эти нейронные сети к 2012 г. Google начала использовать на смартфонах Android. Тот факт, что одну и ту же технологию глубокого обучения смогли применить как для компьютерного зрения, так и для распознавания речи, оказался по-настоящему революционным. Это привлекло внимание к сфере ИИ.

М. Ф.: Удивляет ли вас тот факт, что нейронные сети, с которыми вы много лет назад начали работать, стали центральным элементом проектов в таких крупных компаниях, как Google и Facebook?

И. Б.: Конечно, изначально этого никто не ожидал. В области глубокого обучения был сделан ряд важных, удивительных от-

крытий. Я уже упоминал, что распознавание речи появилось в 2010 г., а о компьютерном зрении стали говорить в 2012 г. Пару лет спустя начался прорыв в сфере машинного перевода, который в 2016 г. привел к появлению сервиса Google Translate. В этом же году началось активное развитие программы AlphaGo. Всего этого мы не ожидали. Помню, как в 2014 г. я просматривал результаты генерации подписей к изображениям и поражался тому, что компьютер смог это сделать. Если бы годом раньше меня спросили, реально ли подобное, я бы ответил «нет».

М. Ф.: Это действительно нечто потрясающее. Конечно, осечки иногда происходят, но в большинстве случаев мы имеем поразительно точный результат.

И. Б.: Осечки неизбежны! Системы пока не обучены на достаточном количестве данных, кроме того, требуется изрядно продвинуться в фундаментальных исследованиях, чтобы они действительно научились распознавать объекты на изображениях и понимать язык. Пока до этого далеко, но ведь даже современного уровня производительности мы изначально не ожидали.

М. Ф.: А как вы пришли к исследованиям в области ИИ?

И. Б.: В юности я активно читал научную фантастику. Подозреваю, что это могло на меня повлиять. Именно оттуда я узнал об ИИ и трех законах робототехники Азимова, и у меня появилось желание изучать физику и математику. А чуть позже мы с братом заинтересовались компьютерами. На сэкономленные деньги мы приобрели компьютер Apple IIe, а затем Atari 800. Программного обеспечения тогда было мало, поэтому мы научились писать программы на языке BASIC.

Я так увлекся программированием, что занялся изучением вычислительной техники, а затем получил ученую степень в области computer science. В 1985 г., во время обучения в магистратуре, я начал читать статьи о первых нейронных сетях, в том числе

работы Джеффри Хинтона. Это было любовью с первого взгляда. Я сразу понял, что хочу работать именно в этой сфере.

М. Ф.: Какой совет вы могли бы дать тем, что хочет заниматься глубоким обучением?

И. Б.: Прыгайте в воду и начинайте плавать. Сейчас информация любого уровня доступна в виде учебников, видео и библиотек с открытым исходным кодом. В сети можно бесплатно прочитать книгу «Глубокое обучение», соавтором которой я являюсь. В ней много информации для новичков. Студенты старших курсов зачастую тренируются, читая научные работы и пытаясь самостоятельно воспроизвести описанные там результаты, затем стараются попасть в лаборатории, проводящие исследования такого рода. Сейчас самое благоприятное время для карьеры в сфере ИИ.

М. Ф.: Из ключевых фигур в сфере глубокого обучения вы единственный, кто занимается только наукой. Большинство по совместительству сотрудничает с различными компаниями. Почему вы выбрали этот путь?

И. Б.: Я всегда высоко ценил научное сообщество, свободу работать на общее благо, делая вещи, которые, как я считаю, могут сильно повлиять на происходящее. Мне нравится работать со студентами, как психологически, так и с точки зрения продуктивности исследований. Уйти работать в индустрию — значит лишиться многого из этих вещей.

Кроме того, я хочу остаться в Монреале, а переход в индустрию означает переезд в Калифорнию или Нью-Йорк. Однажды я подумал, что можно попробовать создать новую Кремниевую долину для ИИ. В результате появился MILA, где проводятся фундаментальные исследования, задающие темп работы над ИИ во всем Монреале. Мы сотрудничаем с научно-исследовательским центром Vector Institute в Торонто и компанией Amii в Эдмонтоне

в рамках канадской стратегии по продвижению ИИ в науке и экономике с пользой для социума.

М. Ф.: Раз уж вы упомянули об экономике, хотелось бы поговорить о рисках в этой сфере. Я много писал о том, что ИИ может привести к новой промышленной революции и потере множества рабочих мест. Как вы относитесь к этой гипотезе? Не преувеличена ли в данном случае угроза?

И. Б.: Нет, она не преувеличена. Непонятно только, когда это произойдет — в ближайшее десятилетие или намного позже. И даже если завтра мы полностью прекратим фундаментальные исследования в области ИИ, те результаты, которых мы уже достигли, позволят кому-то получить социальное и экономическое преимущество за счет простого создания новых товаров и услуг.

Уже собрано огромное количество данных, которые мы пока не используем. Например, в здравоохранении применяется лишь малая доля доступной информации. А ее становится все больше, потому что каждый день оцифровываются новые данные. Производители аппаратных средств совершенствуют процессоры для глубокого обучения, что без сомнения изменит наш мир.

Конечно, прогресс в этой сфере замедляют социальные факторы. Общество не может моментально измениться, даже если технология идет вперед семимильными шагами.

М. Ф.: Реально ли решить проблему безработицы введением безусловного базового дохода?

И. Б.: Я думаю, это может сработать, но сначала нужно избавиться от морального ограничения, согласно которому у неработающего человека дохода быть не должно. Мне такая точка зрения кажется ненормальной. Думаю, нужно ориентироваться на то, что лучше для экономики и для счастья людей. Имеет смысл провести эксперимент, чтобы попытаться найти ответ на эти вопросы.

А единого ответа не будет. Позаботиться о людях, которые в результате новой промышленной революции останутся не у дел, можно разными способами. Мой друг Ян Лекун сказал, что если бы в XIX в. можно было предвидеть последствия промышленной революции, возможно, люди смогли бы избежать множества страданий. Если бы еще тогда, а не в 1950-х мы создали систему социальной защиты, которая сейчас существует в большинстве западных стран, сотни миллионов людей жили бы намного лучше. А ведь для новой революции, скорее всего, потребуется гораздо меньше столетия, и потенциальные негативные последствия могут быть еще сильнее.

Мне кажется, думать об этом нужно уже сейчас. Искать варианты, позволяющие минимизировать нищету и оптимизировать глобальное благополучие. Думаю, выход есть, но мы вряд ли его найдем, если будем держаться за старые ошибки и религиозные убеждения.

М. Ф.: Если все произойдет скоро, это станет еще и политической проблемой.

И. Б.: Поэтому нужно быстро реагировать!

М. Ф.: Совершенно справедливо. А чем еще, кроме влияния на экономику, может грозить ИИ?

И. Б.: Лично я активно выступал против роботов-убийц.

М. Ф.: Я слышал, что вы подписали письмо корейскому университету, который, по слухам, собирался заниматься их разработкой.

И. Б.: Да, и это помогло. Корейский институт передовых технологий (KAIST) сообщил, что они не будут разрабатывать автономные военные системы.

Отдельно я хотел бы коснуться такого важного вопроса, как включение людей в цикл управления. ИИ, который существует

сейчас и может появиться в обозримом будущем, не понимает и не чувствует нормы морали. Разумеется, критерии добра и зла в разных культурах могут отличаться, тем не менее для людей они крайне важны.

Это касается не только роботов-убийц, но и роботов вообще. Представьте себе работу судьи. Для решения таких сложных моральных вопросов нужно понимать человеческую психологию и иметь моральные ценности. Нельзя передавать право принятия судьбоносных решений бездушной машине. Нужны социальные нормы или законы, гарантирующие, что в обозримом будущем компьютеры не получат таких полномочий.

М. Ф.: Здесь я мог бы с вами поспорить. Ваш взгляд на людей и их суждения крайне идеалистический.

И. Б.: Возможно. Но лично я предпочту, чтобы меня судил несовершенный человек, а не машина, не понимающая, что она творит.

М. Ф.: Но представьте себе автономного робота-охранника, который начинает стрелять только в ответ, когда в него попадает пуля. Человеку это недоступно, и потенциально именно такое поведение может спасти другие жизни. Теоретически, если запрограммировать такого робота правильно, он будет избавлен от расовых предрассудков. И в результате он получит преимущество перед человеком. Вы согласны?

И. Б.: Допускаю, что когда-нибудь такое станет возможным. Но вопрос в понимании машиной контекста задачи. Об этом компьютеры не имеют ни малейшего представления.

М. Ф.: Какие еще угрозы может нести ИИ?

И. Б.: Пока это практически не обсуждается, но после происшествий в Facebook и в Cambridge Analytica проблема может выйти на первый план. Она касается рекламы. Применение ИИ с целью воздействия на людей несет опасность для демократии и морально

недопустимо. Общество должно позаботиться о предотвращении подобных явлений.

Например, в Канаде запрещена реклама, направленная на детей. Считается, что манипулировать уязвимыми умами аморально. Разумеется, уязвимы не только дети, иначе реклама бы просто не работала.

Во-вторых, реклама негативно влияет на состояние рынка, потому что за счет нее крупные компании мешают своим малоизвестным конкурентам. Современные технологии на базе ИИ позволяют еще точнее доносить посыл до целевой аудитории. Страшно то, что так людей можно заставить ухудшать собственную жизнь. Я имею в виду, например, политическую рекламу. С инструментами, которые позволяют влиять на людей, следует быть очень осторожными.

М. Ф.: Как вы можете прокомментировать предупреждения Илона Маска и Стивена Хокинга о смертельной угрозе, которую несет суперинтеллект, и о рекурсивном улучшении? Стоит ли об этом беспокоиться в данный момент?

И. Б.: Лично меня эти вопросы не волнуют. Исходя из текущего состояния дел, эти сценарии попросту нереалистичны. Они не совместимы с тем путем, по которому сейчас создается ИИ. Через несколько десятилетий все может измениться, но в настоящий момент — это научная фантастика. По крайней мере, с моей точки зрения. Более того, эти страхи отвлекают от насущных проблем, над которыми мы могли бы работать.

Кроме роботов-убийц и политической рекламы, существует, к примеру, проблема системной предвзятости в данных, ведущая к усилению дискриминации. Правительство и бизнес могут повлиять на это. Поэтому вместо обсуждения рисков, которые могут появиться в долгосрочной перспективе, нужно уделять внимание актуальным угрозам.

М. Ф.: А что вы думаете о работе в этой сфере, которую проводит Китай и другие страны? Например, вы упоминали об ограничениях на автономное оружие, но проблема в том, что некоторые страны могут игнорировать соглашение. Есть ли в данном случае повод для беспокойства?

И. Б.: Как ученый я не считаю это проблемой. Чем больше исследователей во всем мире работает над какой-то темой, тем лучше. Если Китай много инвестирует в исследования в сфере ИИ, это прекрасно; в конце концов, пользоваться результатами мы будем вместе. Хотя меня и пугают мысли о том, что китайское правительство может использовать технологию в военных целях. Системы, умеющие распознавать лица и следить за людьми, позволяют за считаные годы построить общество «Большого Брата». Технически это вполне осуществимо и представляет большую опасность для демократии. Это то, о чем мы должны беспокоиться. Подобное возможно при автократии.

Что же касается гонки вооружений, не нужно смешивать роботов-убийц и применение ИИ в военных целях. Я не считаю, что следует полностью запретить ИИ в армии. Если ИИ будет использован для создания оружия, уничтожающего роботов-убийц, это хорошо. Аморально создание таких роботов, а не применение ИИ военными. Ведь работать можно и над оборонительным оружием.

М. Ф.: То есть вы считаете, что нужен свод правил работы над автономным оружием?

И. Б.: Свод правил требуется везде. По крайней мере, в областях, где применение ИИ будет влиять на общество. Нужно разработать правильные социальные механизмы, которые смогут гарантировать, что ИИ не будет использован во вред.

М. Ф.: И вы думаете, правительство в состоянии заняться этим вопросом?

И. Б.: Доверять решение этого вопроса компаниям точно не следует, потому что их в основном заботит увеличение прибыли. Конечно, они будут пытаться сохранить популярность у пользователей и клиентов, но их действия не совсем прозрачны.

Я думаю, что основную роль тут должно сыграть правительство, точнее даже международное сообщество.

М. Ф.: Считаете ли вы, что выгоды от ИИ в целом перевешивают связанные с ним риски?

И. Б.: Выгоды смогут перевесить риски, если мы будем действовать мудро. Именно поэтому так важно принимать правильные решения. И не хочется, закрыв глаза, мчаться вперед; нужно видеть все подстерегающие нас опасности.

М. Ф.: Где, по вашему мнению, все это должно обсуждаться? В аналитических центрах и университетах? Или требуется политическая дискуссия как на национальном, так и на международном уровне?

И. Б.: Нужна именно политическая дискуссия. В частности, на встрече Большой семерки, куда меня пригласили, был поставлен вопрос: «Какой путь развития ИИ может оказать положительное влияние на экономику и позволит сохранить доверие людей?» Потому что общество обеспокоено. И устранить это беспокойство поможет только открытая дискуссия, в которой смогут участвовать все желающие. Потому что ИИ и связанные с ним проблемы должны быть понятны любому человеку.

> *Как только сильный ИИ «выйдет из детского сада», он превзойдет людей во всех возможных областях и будет обладать куда большей базой знаний, чем любой человек.*

Стюарт Рассел

ПРОФЕССОР ЭЛЕКТРОТЕХНИКИ И COMPUTER SCIENCE. ДИРЕКТОР CENTER FOR INTELLIGENT SYSTEMS ПРИ КАЛИФОРНИЙСКОМ УНИВЕРСИТЕТЕ В БЕРКЛИ

Стюарт Рассел известен как один из ведущих разработчиков ИИ. Является соавтором учебника по ИИ «Искусственный интеллект. Современный подход»[1]*, который в настоящее время используется более чем в 1300 колледжах и университетах в 118 странах. Получил степень бакалавра физики в Уодхэм-колледже Оксфордского университета и докторскую степень в области computer science в Стэнфорде. Занимался исследованиями на различные темы, связанные с ИИ, такие как машинное обучение, представление знаний и компьютерное зрение. Имеет многочисленные награды, в том числе Международной объединенной конференции по ИИ (IJCAI). Является членом Американской ассоциации содействия развитию науки, Ассоциации по продвижению ИИ (AAAI) и Ассоциации вычислительной техники (ACM).*

[1] Рассел С., Норвиг П. Искусственный интеллект. Современный подход. 2-е изд. / Пер. с англ. К. Птицына. — М.; СПб.: Диалектика, 2019. — 1407 с.: ил.

Мартин Форд: Вы написали учебник по ИИ, поэтому мне было бы интересно услышать, как вы определяете некоторые ключевые термины. Что входит в понятие ИИ? Какие проблемы информатики относятся к нему? Как ИИ связан с машинным обучением?

Стюарт Рассел: Я дам вам, скажем так, стандартное определение, которое приведено в нашей книге и в настоящее время общепризнано: «сущность разумна настолько, насколько правильно она поступает». Это означает, что ее действия должны приводить к поставленным целям. Определение относится как к людям, так и к машинам. Если разложить идею правильного поведения на составляющие и исследовать, окажется, что система ИИ должна уметь постигать, видеть, распознавать речь и действовать.

Еще требуется умение видеть суть вещей. Невозможно успешно функционировать в мире, о котором вам ничего не известно. Понять, каким образом мы осознаем различные вещи, помогает такое научное направление, как представление знаний. В его рамках изучаются способы внутреннего хранения данных, с последующей их обработкой алгоритмами формирования рассуждений, такими как алгоритмы автоматического логического вывода и вероятностного вывода.

Машинное обучение всегда было частью науки об ИИ. По сути, это развитие корректного поведения на базе предшествующего опыта.

М. Ф.: Еще дайте, пожалуйста, определения нейронным сетям и глубокому обучению.

С. Р.: Одна из стандартных методик машинного обучения — это обучение с учителем. Системе ИИ дается набор примеров какого-то понятия, снабженных описаниями и метками. Представьте фотографию с подписью, которая указывает, что это изображение лодки, далматинца или чашки с вишнями. Цель обучения состоит в поиске параметра или гипотезы, которые позволят классифицировать изображения в целом. Так мы пытаемся научить ИИ

предсказывать, как могут выглядеть другие изображения тех же объектов.

Гипотезу или параметр можно представить в виде нейронной сети — схемы, состоящей из набора слоев. Входом в нее могут быть значения пикселов на фотографиях далматинцев. В процессе их распространения по схеме на каждом уровне вычисляются новые значения. На выходе из нейронной сети мы получаем распознавание объекта. И мы надеемся, что если подать на вход изображение далматинца, то после прохождения значений всех пикселов через все слои нейронной сети индикатор далматинца будет иметь высокое значение, а индикатор чашки с вишнями низкое. В этом случае можно сказать, что нейронная сеть правильно распознала объект.

М. Ф.: А как заставить нейронную сеть распознавать объекты на изображениях?

С. Р.: Для этого и нужен процесс обучения. Его алгоритмы настраивают весовые коэффициенты всех связей таким образом, чтобы на примерах сеть запоминала правильные ответы. При определенном везении сеть начинает распознавать объекты и на новых, не входящих в обучающий набор изображениях.

Глубокое обучение — это обучение многослойных нейронных сетей. Формально минимального требования к глубине сети не существует, но двух- или трехуровневые сети, как правило, не считаются глубокими. Некоторые сети могут насчитывать более тысячи слоев. В них преобразование, происходящее между входом и выходом, можно представить как композицию более простых преобразований, происходящих на отдельных уровнях. Предполагается, что наличие множества уровней облегчает поиск обобщающих параметров благодаря установлению весовых коэффициентов всех связей.

Мы только подходим к теоретическому пониманию того, в каких случаях и почему глубокое обучение дает верные результаты. По

большому счету все происходящее до сих пор выглядит для нас как магия. Кажется, что изображения, звуковые сигналы и речь, подаваемые на вход глубокой сети, обладают каким-то свойством, помогающим вычленить из них нужный признак. Но пока не ясно, каким.

М. Ф.: Может сложиться впечатление, что ИИ — это синоним глубокого обучения. Это не так?

С. Р.: Приравнивать глубокое обучение к ИИ — ошибка, потому что умение отличать далматинцев от ваз с вишнями — это малая часть требований к эффективному ИИ. Программы AlphaGo и AlphaZero привлекли внимание СМИ к глубокому обучению, но на самом деле это гибрид классического ИИ, который использует метод поиска, с алгоритмом глубокого обучения, который оценивает каждую игровую позицию. Хотя умение отличать хорошую позицию от плохой в го ключевое, программа не смогла бы сыграть на уровне чемпиона мира только в результате глубокого обучения.

По такому же принципу работает система беспилотного автомобиля. На дороге то и дело возникают ситуации, разрешение которых должно происходить по классическим правилам, но в то же время нужно предугадывать возможную реакцию других участников движения, оценивать последствия.

Восприятие — это важный компонент ИИ, который вполне адекватно удается реализовать через глубокое обучение, но для создания системы ИИ требуется множество других способностей различного типа. Особенно это касается действий, растянутых во времени, таких как поездка в отпуск, или сложных — строительство завода. Такие виды деятельности невозможно организовать, имея только систему типа «черный ящик» с глубоким обучением. Иначе алгоритму глубокого обучения нужно будет продемонстрировать все способы, которые когда-либо применялись для строительства. Научится ли система после этого строить заводы? Нет. Во-первых, таких данных не существует, а если бы они и были — нет смысла строить заводы таким образом.

Для строительства нужны специальные знания. Умение планировать. Знание свойств материалов. Чтобы решать долгосрочные и сложные задачи, можно создать системы ИИ, но глубокое обучение тут не поможет.

М. Ф.: Есть ли достижения в сфере ИИ, которые можно считать прорывом?

С. Р.: Хороший вопрос. Дело в том, что многие достижения, о которых активно говорили в СМИ, это не концептуальный прорыв, а всего лишь демонстрация. Вспомните хотя бы победу суперкомпьютера Deep Blue над Каспаровым. По сути, речь шла о демонстрации алгоритмов, разработанных тридцатью годами ранее и постепенно совершенствовавшихся на более мощном оборудовании. Но прорыв заключался в особенностях шахматной программы. В ней интересны и способ прогнозирования, и альфа-бета-алгоритм, сокращающий объем поиска, и некоторые из методов проектирования функций оценки. В итоге, как это часто бывает, СМИ назвали прорывом то, что им не является.

Также и сегодня. Вспомните отчеты о восприятии и распознавании надиктованной речи, заголовки в газетах о точности понимания текста на уровне человека или еще точнее. Но все эти впечатляющие практические результаты — только демонстрация прорывов, произошедших в 1980–1990-х гг.

Сейчас к более старым достижениям прибавлены современные методы проектирования, огромные наборы данных, многоуровневые сети и новейшее оборудование. Есть интерес к ИИ. Но обсуждаются не прорывы.

М. Ф.: Можно ли считать примером прорывной технологии программу AlphaZero от DeepMind?

С. Р.: Это интересная программа. Но нет ничего удивительного в том, что программное обеспечение для игры го смогли использовать для игры в шахматы и сеги на уровне чемпионов мира.

Тот факт, что программа AlphaZero менее чем за сутки научилась играть на сверхчеловеческом уровне в три разные игры, используя одно и то же программное обеспечение, безусловно, вызывает волнение. Но это всего лишь подтверждает, что если вы четко понимаете класс задачи, особенно детерминированной, если есть два игрока, делающих ходы по очереди, а игра идет по известным правилам и за ней можно наблюдать, то решением может стать хорошо спроектированный класс алгоритмов ИИ, позволяющих обучать функции оценки и использовать классические методы управления поиском.

Но для других классов задач, где часть информации скрыта, нужны другие алгоритмические структуры. И программа AlphaZero не научится играть в покер и водить автомобиль. Она не умеет оценивать скрытые факторы. Фигуры на доске для нее — это только фигуры на доске.

М. Ф.: Но ведь в Университете Карнеги — Меллона разработали систему ИИ для игры в покер. Программу Libratus можно считать прорывом в сфере ИИ?

С. Р.: Эта программа — еще один впечатляющий пример гибридного ИИ. Она представляет собой комбинацию алгоритмов, которые существуют уже 10 или 15 лет. Для игр с неполной информацией требуется смешанная стратегия. Ведь если, к примеру, все время блефовать, соперники это быстро поймут. Но без блефа невозможно воровать блайнды, если у вас слабая комбинация карт. Давно известно, что в таких карточных играх нужно постоянно менять свое поведение и оценивать перспективу будущей ставки. ИИ умеет рассчитывать все эти вероятности крайне точно, но только для игры с короткой колодой. Уже примерно 10 лет идет работа над масштабированием расчетов, и есть результат.

Libratus можно назвать впечатляющим современным приложением с ИИ. Но я не уверен, что его методы допускают неограниченное масштабирование. Ведь даже для перехода от первой версии

покера к следующей потребовалось десятилетие. Кроме того, пока непонятно, насколько в реальном мире применимы теории, лежащие в основе игры в покер. Мы не имеем представления о степени рандомизации событий повседневной жизни. Теория игр должна описывать обычную жизнь, но насколько она применима к ИИ, пока неизвестно.

М. Ф.: Когда, по вашим оценкам, автономный транспорт станет общедоступным, и после звонка в Uber приедет пустой автомобиль, который отвезет заказчика в указанное место?

С. Р.: Сроки внедрения беспилотных автомобилей — важный экономический вопрос, ведь в эти проекты инвестируют многие компании.

Но надо упомянуть, что первая беспилотная машина появилась на дорогах общего пользования 30 лет назад! Предложенный Эрнстом Дикманом прототип мог менять ряд, отслеживать положение других транспортных средств на дороге и даже совершать обгоны. Все уперлось в вопросы безопасности. Одно дело успешная демонстрация в течение короткого времени, а совсем другое — система ИИ, которая сможет работать десятилетиями без значительных сбоев. Только во втором случае транспортное средство квалифицируется как безопасное. Пока ничего подобного не существует.

Результаты проводимых сейчас в Калифорнии испытаний показывают, что люди по-прежнему считают необходимым то и дело вмешиваться в управление транспортным средством. Есть более-менее успешные проекты, связанные с беспилотными автомобилями, впечатляющие отчеты показала компания Waymo. Однако пока такие автомобили зависят от погодных условий.

Я считаю, что, если нам повезет, об общедоступности беспилотных автомобилей можно будет говорить где-то через пять лет. Конечно, я не знаю, насколько терпеливо готовы ждать крупные

автомобильные компании. Подозреваю, все они стараются не упустить рыночные возможности.

М. Ф.: Я обычно называю срок 10–15 лет. Ваша оценка не слишком оптимистична?

С. Р.: Да. Но я упомянул о большом везении. Так что в реальности, скорее всего, переход случится намного позднее. Очевидно пока только одно: многие ранние варианты относительно простых архитектур для беспилотных автомобилей в настоящее время забыты, потому что появились новые данные.

Ранние версии автомобилей Google использовали собранную датчиками информацию, которая достаточно хорошо позволяла распознавать другие транспортные средства, разметку, препятствия и пешеходов. Эта система зрения эффективно передавала данные в логической форме, после чего контроллер задавал дальнейшие действия автомобиля в соответствии с заложенными в него логическими правилами. Но в программу каждый день добавлялись новые правила на все случаи жизни. С моей точки зрения, в долгосрочной перспективе архитектура такого типа бесполезна, ведь отсутствие какого-то примера может стать вопросом жизни и смерти.

Мы же не играем в шахматы или го, руководствуясь правилами для конкретных комбинаций фигур. Нигде не написано, что если король стоит на этой клетке, а ладья — на вот той, а еще вон там — ферзь, то делайте вот такой ход. Шахматные программы пишутся по другому принципу. Мы берем правила игры в шахматы и смотрим, к каким последствиям приводят те или иные действия. Так же и в беспилотном автомобиле ИИ должен принимать решения с упреждением.

М. Ф.: Расскажите, что это такое — сильный ИИ и что мешает его создать?

С. Р.: Термин введен недавно. По большому счету это просто напоминание о цели создать ИИ, похожий на интеллект человека.

Поэтому именно сильный ИИ можно называть настоящим ИИ. Этой целью часто пренебрегали ради решения прикладных задач. Потому что решить конкретную задачу, такую как игра в шахматы, намного проще.

Если мы сможем победить ограничения, появится система ИИ, успешно работающая практически в любых обстоятельствах. Ее можно будет попросить как спроектировать новый скоростной катер, так и накрыть стол к ужину. Она сможет выяснить, что не так с вашей собакой, например, прочитав всю доступную информацию о болезнях собак и их лечении. Считается, что именно такие способности отражают универсальность человеческого интеллекта. То есть, говоря о сильном ИИ, мы подразумеваем универсальный ИИ, который в чем-то может нас превзойти. Как только сильный ИИ «выйдет из детского сада», он превзойдет людей во всех возможных областях и будет обладать куда большей базой знаний, чем любой человек.

Но останутся области, в которых машина развиваться не сможет. Это не означает, что сравнивать людей и машины с сильным ИИ не имеет смысла: в долгосрочной перспективе большую важность приобретут их взаимоотношения. Существуют аспекты интеллекта (например, кратковременная память), по которым обезьяна превосходит человека, но будущее горилл и шимпанзе полностью зависит от людей. Когда мы создадим сильный ИИ, то без сомнения столкнемся с проблемой, как избежать судьбы обезьян, как не уступить контроль над собственным будущим.

М. Ф.: Пугающий вопрос. Если концептуальные прорывы порой совершаются за десятилетия их практического применения, есть ли уже признаки скорого появления сильного ИИ?

С. Р.: По моим ощущениям, часть «строительных блоков» для будущего сильного ИИ уже готова. Уже известны и способы представления знаний, и способы рассуждений. Вычислительная логика давно разработана. О представлении логических выкладок в виде

алгоритмов задумывались еще до появления компьютеров. Мы просто пока не знаем, как связать все это с глубоким обучением.

Уже есть такая технология, как вероятностное программирование, объединяющая возможности обучения, логических языков и языков программирования. С математической точки зрения это способ записи вероятностных моделей, в которые можно добавлять данные и получать прогнозы, используя вероятностный вывод.

Моя группа пользуется языком вероятностного моделирования BLOG (название расшифровывается как байесовская логика). Он позволяет записать известную информацию в форме BLOG-модели, к которой можно добавить данные и получить прогноз.

В качестве реального примера можно привести систему мониторинга за соблюдением договора о запрещении ядерных испытаний. В нее записали все, что известно о физике Земли, о распространении и обнаружении сейсмических сигналов, о присутствующих шумах, о расположении станций обнаружения и т. д. Получилась модель, написанная на формальном языке и учитывающая различные неопределенности. Например, мы не знаем, с какой скоростью сигнал будет распространяться в толще Земли. Со станций обнаружения в модель поставляется необработанная сейсмическая информация, а модель отвечает на следующие вопросы: какие сейсмические явления произошли сегодня? Где? Насколько глубоко? Какова их сила? Какие из них могут быть ядерными взрывами? Это активная, довольно хорошо работающая система мониторинга.

Однако невозможно четко смоделировать процесс понимания естественного языка, на котором описывается процесс рассуждений. Если это сделать, то сильный ИИ сможет прочитать учебник по химии, а затем решить экзаменационные задачи, обосновав свое решение, продемонстрировав рассуждения и выводы, ведущие к ответам.

М. Ф.: Или представим систему сильного ИИ, которая читает учебник истории, а затем моделирует современную геополитическую ситуацию, применяя полученные знания в совершенно другой области.

С. Р.: Это хороший пример, который к тому же связан со способностью системы ИИ манипулировать нашим миром в геополитическом или финансовом планах. ИИ уже может подсказывать новые стратегии продвижения товара, помогая компании обойти конкурентов.

Я бы сказал, что реализация способности понимать язык и использовать понятую информацию и есть тот прорыв, которого не хватает для создания сильного ИИ.

Еще не хватает способности к долговременной деятельности. Программа AlphaZero умеет думать на 20, а то и на 30 шагов вперед, но это ничто по сравнению с работой человеческого мозга. Мельчайшие действия человека сопровождаются отправкой сигналов мозга мышцам; даже при наборе одного абзаца текста отправляется несколько десятков миллионов таких сигналов. Соответственно, 20 или 30 шагов дадут сильному ИИ преимущество всего в несколько миллисекунд. AlphaZero бессмысленно использовать для планирования действий робота.

М. Ф.: Как же люди умудряются функционировать, если для этого требуется так много вычислять и решать?

С. Р.: В реальном мире люди и роботы могут действовать только на разных уровнях абстракции. Мы планируем свою жизнь примерно так: «Сегодня днем попытаюсь сделать вот это». А затем абстрактное действие мы разбиваем на более мелкие задачи. Для нас это рутина, но мы не понимаем, как реализовать подобное в системе с ИИ. Поведение человека явно структурировано в виде уровней абстракции. Но откуда взялась иерархия? Как мы создаем и используем такие конструкции? Умея это, машина смогла бы

успешно работать в сложных условиях длительное время. И мы приблизились бы к сильному ИИ.

М. Ф.: А когда, по вашим оценкам, это может произойти?

С. Р.: Подобные вещи никак не связаны с размером наборов данных и мощностью аппаратуры, поэтому количественный прогноз попросту невозможен. Но насколько я помню, 11 сентября 1933 г. Эрнст Резерфорд сказал, что, по его мнению, извлечь из атомов энергию невозможно. На следующее утро, прочитав доклад Резерфорда, Лео Силард разозлился и изобрел ядерную цепную реакцию, запускаемую нейтронами! В итоге сказанное Резерфордом «никогда» 16 часов спустя стало реальностью. Такие примеры внушают оптимизм!

М. Ф.: Вы надеетесь, что сильный ИИ появится при вашей жизни?

С. Р.: Когда на меня начинают так давить, я иногда говорю, что ожидаю его появления при жизни моих детей. На самом деле это просто уход от прямого ответа, потому что к этому времени могут появиться технологии продления жизни.

Благодаря таким компаниям, как Google, Facebook, Baidu, над задачей работает множество очень умных людей. В ИИ вкладываются огромные ресурсы. Темой активно интересуются студенты. Большинство ученых, занимающихся этой темой, считают появление сильного ИИ делом недалекого будущего.

М. Ф.: А как вы думаете, что произойдет после появления сильного ИИ?

С. Р.: В этом случае многие вещи подойдут к своему финишу. Появятся новые измерения интеллекта, и одна за другой будут решаться различные проблемы. Например, сверхчеловеческое мышление можно применить к военной и корпоративной стратегии. Эти инструменты могут появиться у машины раньше, чем умение читать и понимать сложный текст. Первые системы сильного ИИ вряд

ли смогут самостоятельно знакомиться с устройством мира или управлять им. Во многих отношениях они будут догматичными.

М. Ф.: Я хотел бы поговорить о рисках, которые несет ИИ. Ведь, как я знаю, именно на этой теме вы фокусировались в последних работах. Многие считают, что мы приближаемся к новой промышленной революции. Надвигается нечто, что полностью преобразует рынок труда, экономику и прочее. Вы согласны?

С. Р.: Как я уже говорил, невозможно предсказать дату прорыва, после которого ИИ начнет выполнять большую часть человеческой работы. Нельзя составить и однозначный список профессий, которые могут начать исчезать.

В современных дискуссиях и презентациях способности современных технологий сильно завышаются. Кроме того, не осознается сложность интеграции новых систем в существующую схему функционирования корпораций, правительств и т. д.

Разумеется, многие рабочие обязанности сводятся к повторяющимся, рутинным действиям, и имеет смысл для их выполнения заменить человека роботом.

Мне кажется, что в правительствах сейчас думают примерно так: «Нужно начинать подготовку специалистов, которые будут заниматься аналитикой данных и обслуживанием роботов». Но проблема в том, что нам не потребуется миллиард таких специалистов, все ограничится миллионами. Возможно, в маленьких странах, таких как Сингапур или ОАЭ, это жизнеспособная стратегия. Но в большой стране для таких специалистов не будет достаточного количества рабочих мест. Так что в долгосрочной перспективе проблема с безработицей решения не имеет.

Я вижу два варианта развития событий.

В первом случае большинство просто не будет работать, так как будет введен универсальный базовый доход. Ведь после автома-

тизации возрастет продуктивность, появится изобилие товаров и услуг, которое позволит в той или иной форме субсидировать всех неработающих. Человек лишится множества вещей, необходимых для поддержания интереса и стимула к жизни и созиданию.

Во втором варианте будущего, несмотря на то что машины возьмут на себя заботу о выпуске множества товаров и предоставлении базовых услуг, люди займутся вещами, улучшающими качество жизни. Будет цениться умение обучать, вдохновлять на более богатую, интересную, разнообразную и насыщенную жизнь. Учиться ценить литературу и музыку или выживать в дикой местности.

М. Ф.: Думаете, можно направить людей ко второму варианту будущего?

С. Р.: Разумеется, для реализации второго, позитивного варианта потребуется вмешательство извне. Движение в этом направлении нужно начинать прямо сейчас. Строить мир, обеспечивающий эмоциональную устойчивость и воспитывающий конструктивное и позитивное отношение к собственной и к чужой жизни. На данный момент мы не умеем жить таким способом.

Еще я думаю, что нужно коренным образом поменять отношение к науке и тому, что она может дать. На исследования и разработку мобильного телефона потрачены миллиарды долларов. При этом мы почти не пытаемся понять, как вести интересную и полноценную жизнь, как помогать окружающим и в какой момент. Сейчас этим никто не занимается, в этой области нельзя получить научную степень, СМИ практически не пишут об этом, а опубликованные материалы не воспринимаются всерьез.

Я допускаю появление в будущем прекрасно функционирующей экономики, в которой люди, имеющие опыт в разных областях, делятся им с остальными как коучи, преподаватели или психотерапевты.

Это будущее гораздо лучше нашего настоящего; но для перехода к нему нужно переосмыслить систему образования, научную базу,

экономические структуры, определить, по какой схеме будут распределяться доходы. Хотелось бы избежать разделения на богачей, владеющих роботами и системами ИИ, тех, кто им прислуживает, и весь остальной мир, который ничем не занят. С экономической точки зрения это наихудший из возможных сценариев.

М. Ф.: Как в Беркли, так и в Калифорнийском университете в Сан-Франциско вы работали над машинным обучением на базе медицинских данных. Как вы думаете, сможет ли ИИ улучшить ситуацию в области здравоохранения?

С. Р.: Думаю, да. Но мне кажется, что в такой сфере, как медицина, лучше будут работать подходы, базирующиеся на накопленных знаниях и построенных моделях, а не машинное обучение.

Хотя допускаю, что в каких-то областях медицины терабайты маркированных данных дадут хорошие результаты. Активная работа такого рода велась в 1960–1970-х гг., что привело к определенному прогрессу систем ИИ в области медицины. Но современные технологии показали несовершенство имеющихся моделей. По большому счету это модели полностью предсказуемого вымышленного человека.

Вероятностные системы — более обоснованный и целесообразный подход. Они позволяют добавлять к классическим моделям физиологии собираемые в реальном времени данные наблюдений, чтобы поставить диагноз и запланировать лечение.

М. Ф.: А какие риски несет применение ИИ в качестве оружия?

С. Р.: Я думаю, уже началась новая гонка вооружений. В рамках которой, возможно, уже ведется разработка автономного оружия, которому достаточно описания миссии, чтобы самостоятельно идентифицировать и атаковать цель.

Ужаснее всего то, что из автономности логически вытекает масштабируемость. Если для каждой единицы вооружения не требу-

ется человек-оператор, ничто не помешает, например, активировать 10 млн орудий убийства и уничтожить в отдельно взятой стране всех мужчин в возрасте от 12 до 60 лет. Это оружие массового уничтожения.

Применение ядерного оружия — это все-таки переход определенной черты, от которого мы пусть с трудом, но удержались даже в разгар холодной войны. У автономного оружия такого порога нет, к тому же оно легко распространяется. Поэтому, как только будет налажено его производство, доступ к нему получат в числе прочих люди, которые без колебаний пустят его в ход.

М. Ф.: Но ведь в военных целях могут применяться и обычные устройства. Достаточно приобрести на сайте Amazon беспилотный летательный аппарат и добавить к нему оружие…

С. Р.: Современные беспилотные летательные аппараты все-таки управляются с земли. Конечно, к такому аппарату можно прикрепить небольшую бомбу и доставить ее куда нужно, но сам по себе он этого сделать не сможет. Кроме того, для запуска 10 млн дронов потребуется столько же людей-операторов. Гипотетически можно представить, что кто-то подготовит целую армию для управления оснащенными бомбами дронами, но такие вещи попадают под санкции международных контролирующих органов. А вот возможности контролировать автономное оружие пока не существует.

М. Ф.: Но разве невозможно в домашних условиях разработать автономную систему управления и развернуть ее на дронах? Контролируются ли как-нибудь подобные вещи?

С. Р.: Да, для управления оснащенными оружием дронами можно использовать нечто похожее на программное обеспечение, которое управляет беспилотным автомобилем. В результате получится самодельное автономное оружие. Возможно, по этому поводу со временем будет заключен некий регулирующий договор, придумают механизм проверки с привлечением производителей

дронов и чипов для беспилотных автомобилей. На заметку будут брать всех, кто заказывает подобные комплектующие в больших количествах. Разумеется, плохие вещи все равно будут делаться, но редко. А в небольших количествах автономное оружие не обладает особыми преимуществами перед обычным.

Кроме того, с применением ИИ в военных целях связаны и другие риски. Например, машины могут неправильно истолковать какой-то сигнал. Нельзя сбрасывать со счетов и вероятность взлома, после чего автономное оружие, которое вы считаете своей надежной защитой, обернется против вас.

М. Ф.: Эти пугающие сценарии вы показали в короткометражном фильме *Slaughterbots* («Роботы-убийцы»).

С. Р.: В письменной форме мы не смогли достучаться до широких масс. Люди продолжали считать автономное оружие научной фантастикой.

М. Ф.: В 2014 г. вместе со Стивеном Хокингом, Максом Тегмарком и Фрэнком Вильчеком вы опубликовали письмо[1], в котором говорилось о недостаточно серьезном отношении к рискам, связанным с ИИ. Среди авторов вы были единственным специалистом в области computer science. Как появилась идея этого письма?

С. Р.: Все началось с того, что Национальное общественное радио попросило меня поделиться мыслями по поводу фильма «Превосходство». Фильм совсем недавно появился в прокате, поэтому я пошел в кино, понятия не имея о том, что именно увижу.

Сначала там показали отдел информатики в Беркли. Профессор в исполнении Джонни Деппа рассказывал об ИИ, как вдруг его убил представитель радикальной антитехнологической груп-

[1] Оригинал письма доступен по адресу https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html

пировки. В этот момент я невольно сжался, ведь на месте этого профессора в то время вполне мог оказаться и я. Перед смертью ученого его жена и лучший друг успели загрузить его мозг в большой квантовый компьютер. Он стал сверхинтеллектуальной сущностью, которая смогла быстро развить различные новые технологии и попыталась захватить мир.

В итоге появилось письмо, которое на первый взгляд представляло собой рецензию на фильм, но на самом деле несло следующую идею: «Любые создаваемые машины могут оказать сильное влияние на реальный мир, и это может стать серьезной проблемой, потому что речь идет о передаче контроля над нашим будущим объектам, которые не являются людьми». Ведь именно интеллект дает нам способность контролировать мир.

М. Ф.: Но многие видные ученые, занимающиеся ИИ, не считают это проблемой…

С. Р.: Да, я слышал много раз, что связанные с ИИ проблемы надуманны. Но пока аргументы сторонников такой точки зрения не показались мне заслуживающими внимания. Например, утверждать, что механизм с ИИ можно просто отключить, все равно что считать, что для победы в го над компьютером AlphaZero достаточно правильно двигать свои камни.

Отрицание проблемы ИИ я воспринимаю как защитную реакцию. Чтобы перестать бояться угрозы, люди убеждают себя, что ее не существует. По крайней мере, эта теория объясняет, почему даже очень информированные люди пытаются отрицать наличие проблемы.

Распространяется теория и на тех представителей ИИ-сообщества, которые не верят в возможность появления сильного ИИ. Здесь ситуация еще забавнее, потому что у нас за плечами 60 лет работы над темой, в успех которой мало кто верил. Но вещи, которые, если верить скептикам, были невозможными, вопло-

тились в реальности. Например, компьютер победил чемпиона мира по шахматам.

Я верю, что люди, которые за историю своего существования сделали множество открытий и даже научились пользоваться атомной энергией, смогут преодолеть препятствия и создадут интеллект, достаточно сильный для того, чтобы возникла угроза его выхода из-под контроля. Благоразумно начать думать над тем, как убрать эту угрозу еще на стадии проектирования систем ИИ.

М. Ф.: А как это можно сделать?

С. Р.: На мой взгляд, мы не совсем корректно определяем понятие «интеллект». В целом, это рациональное поведение. Если создать машины с аналогичной способностью осуществлять свои намерения, появятся очень умные существа, но со своими целями. И достигать они будут именно их!

Первоначально ИИ был задуман как оптимизатор: то есть искатель оптимальных путей достижения человеческой цели. И это работало, но только потому, что созданные машины не были по-настоящему интеллектуальными и функционировали в ограниченном пространстве. После выхода ИИ в реальный мир все может пойти не по плану. Мы уже видели обвал торговых рынков после появления торговых ботов.

Мне кажется правильным, когда машины, помогающие в достижении наших целей, не знают, в чем эти цели заключаются. Нужна неопределенность относительно характера наших целей. Именно она даст желанную гарантию безопасности.

Попробую продемонстрировать это на примере. Если цель машины приносить кофе, при достаточном интеллекте она сможет понять, что в выключенном состоянии она не сможет этого сделать. Тогда она предпримет шаги для предотвращения выключения. Я считаю, что необходимо заложить в машину цель избегать нежелательных для человека действий. Такое видение ИИ можно опи-

сать математически, показав, что гарантия безопасности (то есть в данном случае мотив, под влиянием которого машина позволит себя выключать) напрямую связана с отсутствием информации о целях человека.

Конечно, мое видение ИИ несколько отличается от общепринятого. Но именно так можно построить систему ИИ, более совершенную с точки зрения безопасности и контроля.

М. Ф.: Раз уж речь зашла о безопасности и контроле, поясните, пожалуйста, насколько серьезно следует относиться к гонке вооружений в других странах, особенно в Китае.

С. Р.: Ник Бостром и другие ученые выражали свое беспокойство тем, что, если правительство сочтет стратегическое доминирование в области ИИ важным компонентом национальной безопасности и экономического лидерства, то акцент будет делаться на его максимально быстрое развитие. Забота о степени контролируемости таких систем в этом случае отойдет на второй план.

Эти опасения кажутся мне не лишенными основания. С другой стороны, поскольку мы работаем над системами ИИ, пригодными для использования в реальном мире, у нас есть экономический стимул сделать так, чтобы они оставались под контролем.

Для примера представим продукт, который может появиться в ближайшем будущем: обладающий интеллектом личный помощник, фиксирующий ваши действия, разговоры, отношения и т. п. Фактически это личный секретарь. Вы же понимаете, что систему, которая не разбирается в человеческих предпочтениях и действует небезопасными способами, просто никто не купит. Система, не понимающая нюансов человеческой жизни, может забронировать номер в гостинице за 20 000 долларов в сутки или ответит отказом на предложение встретиться с президентом, потому что на это время вы уже записаны к стоматологу. Кому нужен такой робот? Это станет концом индустрии. Вспомните, что случилось

с атомной промышленностью после Чернобыля и Фукусимы. Если мы не сможем решить проблему контроля, область ИИ умрет.

М. Ф.: Но в целом вы оптимист? Думаете, что все получится?

С. Р.: Да, меня можно назвать оптимистом. Мы только начинаем подходить к проблеме управления, но первые попытки кажутся вполне продуктивными. Поэтому я верю, что существует путь развития, который позволит создать доказанно полезные системы ИИ.

Разумеется, есть риск, что даже если мы решим проблему контроля и построим полезные ИИ-системы, кто-то другой может пойти по пути наращивания возможностей ИИ без учета аспектов, связанных с безопасностью. Эдакий доктор Зло, который создаст разрушительную систему ИИ или будет злоупотреблять контролируемым ИИ. В этих сценариях мы постепенно превратимся в ослабленное общество, в котором так много знаний перенесено в машины и так много решений отдано им на откуп, что вернуть контроль над ситуацией мы уже не сможем и потеряем свободу воли.

Такое будущее изображено в фильме «ВАЛЛ-И», где люди отправляются в космос в попытке спастись, а на кораблях их обслуживают машины. Постепенно люди становятся слабее, ленивее и глупее. О риске оказаться в таком будущем тоже следует помнить.

Но я оптимистично надеюсь на будущее, в котором системы ИИ смогут напоминать людям: «Не используйте нас. Получайте знания и навыки самостоятельно. Сохраняйте свои возможности и способности, распространяйте цивилизацию через людей, а не через машины».

> *В 1980-х гг., когда вокруг ИИ в целом и метода обратного распространения в частности была большая шумиха, люди ждали чудес и не дождались. Но сегодня чудеса уже происходят, а на шумиху не стоит обращать внимание.*

Джеффри Хинтон

ПРОФЕССОР COMPUTER SCIENCE УНИВЕРСИТЕТА ТОРОНТО, РУКОВОДИТЕЛЬ ИНЖЕНЕРНО-ТЕХНИЧЕСКОЙ СЛУЖБЫ КОМПАНИИ GOOGLE, ГЛАВНЫЙ КОНСУЛЬТАНТ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОГО ЦЕНТРА VECTOR INSTITUTE

Джеффри Хинтона иногда называют крестным отцом глубокого обучения. Именно он стоял у истоков таких ключевых технологий, как метод обратного распространения, машина Больцмана и капсульная нейронная сеть.

Джеффри — член Лондонского королевского общества, а также иностранный член Национальной инженерной академии США и Американской академии искусств и наук. Обладатель премии Румельхарта, награды IJCAI за исследовательские достижения, канадской национальной премии Killam Prize в области разработки, награды имени Фрэнка Розенблатта, золотой медали Джеймса Клерка Максвелла от Института инженеров электротехники и электроники (IEEE), награды C & C от японской корпорации NEC, премии от компании BBVA и высшей награды Канады в области науки и техники — золотой медали Герхарда Херцберга.

Мартин Форд: Широкую известность вам принесла работа над методом обратного распространения ошибки. Расскажите, пожалуйста, что это за метод?

Джеффри Хинтон: Проще всего объяснить от обратного. Считается, что существует четкий алгоритм обучения нейронных сетей. Есть сеть, состоящая из слоев нейронов, снизу — вход в нее, сверху — выход. Все связи между нейронами имеют свой вес. Каждый нейрон смотрит на нейроны нижнего слоя и умножает их активность на вес связей, затем складывает все это и выдает результат. Регулируя вес связей, можно заставить сеть выполнять нужные вам операции, например находить кота на изображении и добавлять соответствующую метку.

Но как с помощью регулировки весов добиться от сети нужного результата? Существует простой и эффективный, но невероятно медленный алгоритм. Всем связям присваиваются случайные веса, сети демонстрируются примеры, и вы смотрите, что получается. Затем один из этих весов немного меняется, и демонстрируется другой набор примеров. Если результат работы сети стал лучше, внесенные изменения сохраняются. В противном случае вес возвращается к исходному значению или меняется в противоположном направлении. Затем эта операция проделывается с весом следующей связи и т. д.

В результате работа сети оценивается для каждого веса, причем на ряде примеров. Каждый вес должен обновляться несколько раз. Это медленно, но результат гарантирован.

Метод обратного распространения ошибки, по сути, позволяет получить такой же результат намного быстрее. Скорость его работы зависит от количества связей. Для сети с миллиардом связей метод обратного распространения сработает в миллиард раз быстрее, чем описанный выше алгоритм.

Прямой алгоритм имитирует процесс эволюции, ведь то, как заложенная в генах информация реализуется в конкретном индивиде, зависит от среды, в которой он находится. По генотипу

невозможно точно предсказать, как будет выглядеть фенотип или насколько он будет успешным, потому что на это влияет множество внешних факторов.

Поскольку корректность результатов определяется только весами, которые нам известны, процесс прохождения данных можно контролировать с помощью метода обратного распространения. Его суть состоит в передаче сигналов ошибки от выхода к входу. В процессе их прохождения вычисляется, как следует поменять вес каждой связи, чтобы улучшить выводимый результат.

Вместо того чтобы *измерять* эффект от внесенных изменений, метод обратного распространения ошибки *вычисляет*, что получится после внесения изменений, причем для всех весов одновременно. Настройка весов выполняется быстро: сети предоставляется сразу несколько примеров, рассчитывается разность между требуемым выходом и результатами сети, после чего эта информация передается в обратном направлении. Процесс выполняется несколько раз, но все равно работает быстрее эволюционного алгоритма.

М. Ф.: Метод обратного распространения ошибки изобрел Дэвид Румельхарт, а вы развили его?

Дж. Х.: Версии этого метода предлагались еще до Румельхарта. В основном к этой идее приходили независимо друг от друга, поэтому меня всегда смущает, когда в СМИ меня называют автором этого метода. Я главным образом продемонстрировал, как использовать этот метод для изучения распределенных представлений.

В 1981 г. после получения докторской степени я начал работать в городе Сан-Диего, штат Калифорния. Идею метода обратного распространения ошибки предложил Дэвид Румельхарт, а мы с Рональдом Уильямсом помогли в поиске правильных формулировок. Ничего впечатляющего с этим методом мы тогда не сделали. Не было и никаких публикаций. После этого я отправился в Университет Карнеги — Меллона, где работал над машиной Больцмана.

Эта идея казалась более интересной, хотя она и не сработала. В 1984 г. я вернулся в Сан-Диего, чтобы сравнить метод обратного распространения с машиной Больцмана. Оказалось, что он дает более убедительные результаты, поэтому я снова начал общаться с Дэвидом Румельхартом.

Но по-настоящему меня восхитила возможность на примере формирования генеалогического древа применить метод обратного распространения к изучению распределенных представлений. На вход подавалось два слова, а возвращалось третье, связанное с обоими. То есть нейросеть как бы улавливала значения слов.

Например, если мать Шарлотты зовут Виктория, то корректным выводом для слов *Шарлотта* и *мать* было *Виктория*. А для слов *Шарлотта* и *отец* корректным был ответ *Джеймс*. Если взять генеалогическое древо, в котором нет разводов, то стандартный ИИ, используя свои знания о семейных отношениях, может сделать вывод, что Виктория — супруга Джеймса. К такому же выводу может прийти нейронная сеть, причем не пользуясь логическими правилами, а просто изучив множество признаков каждого человека. В этом случае Виктория и Шарлотта — это наборы отдельных признаков. Результат взаимодействия двух векторов дает признаки корректного ответа. Потрясает, как сеть изучает векторы признаков и распределенные представления для разных слов.

В 1986 г. мы описали это в статье для журнала *Nature*. Тема сильно заинтересовала одного из рецензентов. Как психолог, он понимал, что алгоритм, обучающий представлению о вещах, станет огромным прорывом. Так что мой вклад заключается не в открытии алгоритма обратного распространения, а в том, что я смог показать, как этот метод может применяться для обучения распределенным представлениям. Именно это оказалось интересно психологам и, в конечном итоге, людям, которые занимались вопросами ИИ.

В начале 1990-х Иошуа Бенджио перенес этот метод на более быстрые компьютеры. Он применил нейронную сеть к естествен-

ному языку. Сеть брала из текста несколько слов в качестве контекста и могла предсказать следующее слово. Ян Лекун, который в это время занимался компьютерным зрением, показал, что метод обратного распространения хорошо обучает фильтры обработки визуального входа. Это не стало особым открытием, так как примерно такие же вещи делает человеческий мозг. А вот то, что метод обратного распространения позволил машине уловить значения слов и синтаксис, стало большим прорывом.

М. Ф.: Правильно ли я понимаю, что в то время работа с нейронными сетями еще не была основным направлением в исследованиях ИИ?

Дж. Х.: До некоторой степени да, но тут нужно отдельно рассматривать ИИ и машинное обучение, с одной стороны, и психологию — с другой. В 1986 г., когда метод обратного распространения стал популярным, им заинтересовались психологи. Это был устойчивый интерес, хотя алгоритм не копировал происходящие в мозге процессы. А в конце 1980-х гг. Ян Лекун получил впечатляющие результаты по распознаванию рукописных цифр. Метод обратного распространения хорошо себя показал и в других областях, таких как контроль мошенничества с кредитными картами. Но ожидания тех, кто считал, что теперь нам будут доступны настоящие чудеса, не оправдались.

В начале 1990-х гг. оказалось, что на небольших наборах данных лучше себя показывают другие методы машинного обучения. Например, метод опорных векторов с меньшими усилиями распознавал рукописные цифры. И интерес к обратному распространению затух.

Идея метода обратного распространения состояла в обучении множества слоев, но обучить удалось только не очень глубокие сети. С точки зрения специалистов по статистике и ИИ мы были мечтателями, которые надеялись получить информацию обо всех весах только по входным и выходным данным. На тот момент нам не хватало знаний, чтобы заставить все это работать.

До 2012 г. большинство специалистов по компьютерному зрению считали все это сумасбродством, хотя системы Яна Лекуна иногда работали лучше, чем их собственные. Ян написал статью, но ее не приняли, так как считалось, что этот способ не даст результатов. Даже в мире науки альтернативные подходы отвергаются.

Но внезапно крупный конкурс выиграли двое моих учеников. Они применили комбинацию методов, разработанных в лаборатории Лекуна, и наших собственных техник и получили в два раза меньше ошибок, чем лучшие системы компьютерного зрения.

М. Ф.: Речь идет о проекте ImageNet?

Дж. Х.: Да. Там случилось то, что периодически происходит в науке. Метод, который привыкли считать полной бессмыслицей, превзошел метод, в который все верили. За следующие два года все переключились на сверточные нейронные сети. Сейчас никто даже не думает о классификации объектов без использования нейронной сети.

М. Ф.: То есть в 2012 г. наступил переломный момент в отношении глубокого обучения?

Дж. Х.: Это был переломный момент для компьютерного зрения. В сфере распознавания речи он случился раньше. В 2009 г. два аспиранта из Торонто показали, что глубокое обучение позволяет улучшить распознавание речи. Они стали стажерами в IBM и Microsoft, а другой мой студент принес эту систему в Google.

М. Ф.: Если почитать современную прессу, создается впечатление, что нейронные сети и глубокое обучение — это эквивалент ИИ.

Дж. Х.: Долгое время ИИ считался системой, запрограммированной на определенные правила обработки символьных строк. В этом заключался интеллект. Оставалось уточнить, как выглядят эти правила и строки. Затем появились нейронные сети. Это была попытка смоделировать разум по образцу человеческого мозга.

Обратите внимание, что ИИ в том виде, как его изначально понимали, не имел отношения к обучению. В 1970-х гг. упор делался на определение правил и выбор символических выражений. Считалось, что рано думать об обучении. Тех же, кто занимался нейронными сетями, интересовали вопросы обучения, восприятия и управления движением. Они считали, что с точки зрения эволюции способность рассуждать логически появляется на поздних стадиях развития.

Сейчас в промышленности и правительстве термин «ИИ» используют для обозначения глубокого обучения, что приводит к парадоксальным ситуациям. Например, в Торонто было щедро профинансировано создание научно-исследовательского центра Vector Institute, который занимается фундаментальными исследованиями в области глубокого обучения. Разумеется, на финансирование претендуют многие, и один университет заявил, что у них больше специалистов по ИИ, чем в Торонто, приведя в качестве доказательства количество ссылок на их работы. При этом они занимаются классическим ИИ. Мне кажется, что лучше вообще не пользоваться этим термином, чтобы не давать поводов для подобной путаницы.

М. Ф.: Вы считаете, что термин «искусственный интеллект» должен быть связан только с нейронными сетями?

Дж. Х.: Я думаю, общая идея ИИ состоит в создании небиологических интеллектуальных систем. Хотя до сих пор у многих, особенно у старшего поколения ученых, сохраняется наивное представление об ИИ как о системе, умеющей манипулировать символическими выражениями. Но даже наборы символов на входе и выходе не означают, что внутри мы имеем дело с символьными строками. Там находятся векторы нейронной активности.

М. Ф.: В конце 2017 г. вы сказали в интервью[1], что алгоритм обратного распространения ошибки нужно отбросить, все начать с нуля. Что вы имели в виду?

[1] https://www.axios.com/artificial-intelligence-pioneer-says-we-need-to-start-over-1513305524-f619efbd-9db0-4947-a9b2-7a4c310a28fe.html

Дж. Х.: В данном случае неверно передан контекст беседы. Я поднимал вопрос о том, насколько метод обратного распространения может помочь в понимании принципов работы мозга. Есть основания полагать, что в мозге ничего подобного не происходит. Но это не означает, что метод не нужно применять при моделировании искусственных систем.

М. Ф.: И возможно, результаты будут улучшаться?

Дж. Х.: Мы работаем над совершенствованием глубокого обучения. Я допускаю появление других алгоритмов, действующих альтернативными способами. Но даже их появление не значит, что следует отказаться от метода обратного распространения.

М. Ф.: А почему вас заинтересовали ИИ и нейронные сети?

Дж. Х.: Мой школьный друг Инман Харви был очень хорошим математиком. И однажды заинтересовался идеей, что мозг может работать как голограмма.

М. Ф.: Голограмма как трехмерное представление?

Дж. Х.: Вы помните, что если разрезать голограмму, получится не две половины изображения, а нечеткое изображение всей сцены в каждой половине? Такой способ распределения информации отличается от того, к чему мы привыкли. Отрезав кусок фотографии, вы просто потеряете информацию о том, что было на этом фрагменте.

Инмана заинтересовал тот факт, что таким же способом может работать память человека, в то время как каждый отдельный нейрон не отвечает за ее сохранение. Он предположил, что информация в мозге хранится благодаря настройке силы связей между нейронами, и что, по сути, это распределенное представление, то есть попытка изобразить какие-то вещи, которым соответствует активность в ряде нейронов, но каждый нейрон участвует в представлении разных вещей. При таком подходе исключено однозначное соответствие между нейронами и понятиями. Это

первое, что меня заинтересовало. Кроме того, было интересно, каким образом мозг может учиться чему-то, регулируя силу соединений между нейронами.

М. Ф.: И это в старших классах школы? Потрясающе! А как поменялись ваши интересы после поступления в университет?

Дж. Х.: Там я начал изучать физиологию. И узнал, как нейроны распространяют потенциалы действия. Помогли эксперименты на аксоне гигантского кальмара. Оказалось, что именно так работает мозг. Но я разочаровался, узнав, что не существует модели процесса представления или изучения различных вещей.

Я переключился на психологию, думая, что здесь мне скажут, как работает мозг. Но в то время в Кембридже основной упор делался на бихевиоризм, и изучение психологии сводилось к крысам в лабиринтах. Конечно, занимались и когнитивной психологией, но это не имело отношения к вычислительным методам, поэтому мне казалось, что мы никогда не узнаем, как же работает мозг.

В рамках проекта по изучению развития детей я смотрел, как в возрасте от 2 до 5 лет меняется восприятие различных свойств. Оказалось, что совсем маленькие дети в основном интересуются цветом и текстурой, а по мере взросления их начинает интересовать форма. Детям демонстрировались три объекта, один из которых отличался от двух других. Например, два желтых круга и один красный. Мне нужно было научить детей убирать лишний объект.

В следующих заданиях лишним был объект другой формы. После этого детям предложили желтый треугольник, желтый круг и красный круг. Предполагалось, что если детей больше интересует цвет, они уберут красный круг, а если форма — лишним окажется желтый треугольник. Один пятилетний ребенок указал на красный круг и сказал: «Он покрашен в неправильный цвет».

Я работал над подтверждением крайне примитивной модели, которая всего лишь утверждала, что фокус смещается от цвета

к форме, но никак не объясняла почему. Ребенок же обработал информацию, которую ему давали в виде обучающих примеров, и, столкнувшись с набором, допускавшим удаление более чем одного объекта, решил, что, возможно, что-то покрашено не в тот цвет.

Тестируемая модель не была адаптирована к уровню сложности изучаемой системы, и других адекватных моделей не было. Передо мной была интеллектуальная система, умеющая обрабатывать информацию и делать выводы о причинах происходящего. И я прекратил ее исследовать таким неэффективным способом.

М. Ф.: После этого вы начали заниматься ИИ?

Дж. Х.: Нет, я стал плотником. И некоторое время наслаждался этим, пока не увидел работу настоящего мастера. Это меня так расстроило, что я решил вернуться в науку.

М. Ф.: Если вспомнить, что было дальше, наверное, к лучшему, что вам не удалось стать по-настоящему квалифицированным плотником!

Дж. Х.: Дальше я принял участие в проекте по изучению развития речевых навыков у детей. Хотелось понять, каким образом на них влияет принадлежность к определенному социальному классу. Мне нужно было создать вопросник для оценки отношения матерей. Первое же интервью меня ошарашило. Женщина из бедного пригорода Бристоля на вопрос, как она относится к тому, что ребенок разговаривает, ответила: «Если он начинает разговаривать, мы его бьем». На этом моя карьера социального психолога завершилась.

Я поступил в аспирантуру Университета Эдинбурга по специальности ИИ. Моим руководителем был Хью Кристофер Лонге-Хиггинс, который изначально занимался химией в Кембридже, а затем переключился на изучение ИИ. Он верил в компьютерное моделирование, поэтому я попросился к нему. К сожалению, у него как

раз начался переломный период. Он решил, что нейронные модели ничего не дают, а к пониманию интеллекта нужно идти через язык.

Напомню, что как раз тогда появились модели систем, которые могли определять расположение блоков. В них применялась обработка символов. Американский профессор Терри Виноград показал, как заставить компьютер в определенной степени понимать язык, отвечать на вопросы и фактически следовать командам. Можно было сказать: «Покажи блок, который находится в синем ящике на вершине красного куба», и получить желаемое. Лонге-Хиггинса это сильно впечатлило, и он решил над этим работать, я же предпочел продолжить изучение нейронных сетей.

Я считаю Кристофера замечательным ученым, хотя мы и не смогли прийти к согласию в работе. Он поддерживал меня, даже когда я отказывался делать то, что он говорил. В конце концов, я защитил диссертацию по нейронным сетям, хотя в то время они, по общему мнению, были полной чепухой.

М. Ф.: Примерно в то время Марвин Минский и Сеймур Пейперт написали свою книгу «Персептроны»[1].

Дж. Х.: Нет, защитился я в начале 1970-х гг., а книга Минского и Пейперта вышла в конце 1960-х гг. Практически все, кто занимался ИИ, думали, что нейронным сетям конец. Считалось, что интеллект обусловлен программно, и нужно просто понять, какие программы использует мозг.

Что интересно, были и сторонники логического подхода, которые верили в парадигму нейронных сетей. Например, Джон фон Нейман и Алан Тьюринг полагали, что большие сети нейронов способны помочь в изучении интеллекта. Но доминировал подход, в котором одни строки символов формировали другие. Казалось, что модель логических рассуждений должна выглядеть именно

[1] Минский М., Пейперт С. Персептроны / Пер. с англ. Г. Гимельфарба и В. Шарыпанова. — М.: Мир, 1971. — 261 с.: ил.

так. Мало кто верил, что можно понять интеллект, глядя на то, как реализован мозг. Было убеждение, что интеллект должен рассматриваться сам по себе.

М. Ф.: В прессе много писали, что успехи глубокого обучения преувеличены и следует сокращать инвестиции в эту сферу. Мне даже встречалось выражение «зима ИИ». Исследования зашли в тупик?

Дж. Х.: В 1980-х гг., когда вокруг ИИ в целом и метода обратного распространения в частности была большая шумиха, люди ждали чудес и не дождались. Но сегодня чудеса уже происходят, а на шумиху не стоит обращать внимание. Наши мобильные телефоны распознают речь, компьютеры узнают объекты на фотографиях, Google выполняет машинный перевод. Искусственно созданный ажиотаж означает раздачу обещаний, которые никто не собирается выполнять. Но сейчас мы уже многого достигли и не собираемся на этом останавливаться.

В интернете я иногда вижу рекламу, где говорится, что это индустрия стоимостью 19,9 триллиона долларов. Кажется, что это много и похоже на обман, но идея о том, что это сфера с многомиллиардными доходами, явно не просто шумиха, потому что несколько человек вложили в нее миллиарды долларов, и это сработало для них.

М. Ф.: Во что лучше всего инвестировать для продвижения вперед? Ведь все еще остались люди, которые верят в условный ИИ и считают, что оптимально совмещать глубокое обучение с более традиционными подходами. Как вы относитесь к этой точке зрения?

Дж. Х.: Я считаю, что работа мозга — это взаимодействие больших векторов нейронной активности. И именно по такому принципу будет работать ИИ. Разумеется, понять механику рассуждений тоже важно, но, мне кажется, мы сможем это сделать уже после того, как реализуем все остальное.

Я не верю в гибридные системы. Был пример в автомобильной промышленности, когда электродвигатель стали использовать для впрыскивания бензина. Точно так же сторонники условного ИИ, признавая все преимущества глубокого обучения, хотят использовать его как слугу, который обеспечит функционирование символических рассуждений. Это попытка остаться на старых позициях, невзирая на меняющиеся обстоятельства.

М. Ф.: Мне вспоминается ваш последний проект, связанный с так называемыми *капсулами*, имитирующими, как я понимаю, колонки кортекса в мозгу. Считаете ли вы, что результаты изучения мозга следует включать в работу с нейронными сетями?

Дж. Х.: Капсулы появились как сложная и умозрительная комбинация полудюжины разных идей. Об успехах этой концепции пока говорить рано, но ее основой действительно послужила структура человеческого мозга. Думать о работе с нейронными сетями на базе достижений нейробиологии несколько наивно. Ученые пытаются понять базовые принципы функционирования мозга, но при этом есть еще и многочисленные детали, которые по-разному выглядят на разном аппаратном обеспечении. Графические процессоры (GPU) мало напоминают мозг, но это не мешает нам искать общие принципы работы мозга и нейросетей. Например, в обоих случаях бóльшая часть знаний возникает благодаря обучению, а не механическому запоминанию фактов.

Для условного ИИ нужна огромная база фактов и правила их связи друг с другом. Перенести какое-то знание из одного мозга в другой невозможно. В голове множество параметров, то есть весов связей между отдельными нейронами, которые не передадутся. Если один человек расскажет, каким образом что-то работает, другой может повторить эти действия, но не более. И это хорошо, потому что у каждого своя нейронная сеть.

М. Ф.: Правда ли, что глубокое обучение в основном происходит на базе маркированных данных, то есть речь идет о так называе-

мом обучении с учителем, а задача обучения без учителя еще не решена?

Дж. Х.: Это не совсем так. Существует большая зависимость от размеченных данных, но вопрос в том, что считать такими данными. Например, если я попрошу вас предсказать следующее слово в большой текстовой строке, меткой правильного ответа будет слово, которое фактически там фигурирует. Дополнительные метки тут не нужны. А если нужно обучиться распознавать на картинках кошек, потребуется метка «кошка», которая не является частью изображения. Эти метки нужно будет добавлять вручную. По идее первая задача тоже относится к обучению с учителем, потому что мы имеем дело с метками. Но это своего рода промежуточный вариант между маркированными и немаркированными данными.

М. Ф.: Дети по большей части учатся без учителя.

Дж. Х.: Исходя из того, что я говорил выше, ребенок воспринимает окружающую среду, пытаясь предсказать, что будет дальше. Затем, когда что-то происходит, событие помечается как правильно или неправильно угаданное. К сожалению, в случае прогнозирования сложно понять, какой из вариантов, «с учителем» или «без учителя», применяется. Задача, в которой по набору изображений нужно предсказать, какое будет следующим, не попадает ни в одну из двух категорий.

М. Ф.: То есть одно из основных препятствий сейчас — отсутствие общей формы обучения?

Дж. Х.: Да. Но я утверждаю, что решить задачу предсказания могут алгоритмы обучения с учителем.

М. Ф.: А как бы вы определили сильный ИИ?

Дж. Х.: Меня вполне устраивает официальное определение, утверждающее, что это интеллект, сравнимый с человеческим, но

многие думают, что появятся отдельные модули ИИ, которые будут становиться все умнее и умнее. Они не понимают, что нейронные сети в чем-то превосходят людей, а в чем-то сильно уступают им. Например, ИИ-система может лучше интерпретировать медицинские изображения, но ей сложно рассуждать о них.

Представляя отдельные ИИ-системы, люди игнорируют социальный аспект. Для ИИ требуются вычислительные ресурсы, поэтому создание сильного ИИ приведет к появлению сообществ интеллектуальных систем. Ведь сообществу доступно гораздо больше данных, чем отдельной системе. Лучше распределить данные по множеству систем и заставить их общаться друг с другом.

М. Ф.: И все сведется к появлению связанного интеллекта?

Дж. Х.: Но ведь в мире людей все обстоит точно так же. Мы извлекаем знания из источников, составленных для нас другими людьми. А опыт обучения позволяет прийти к аналогичному пониманию.

М. Ф.: Реально ли создать сильный ИИ?

Дж. Х.: В компании OpenAI уже есть система, играющая в сложные компьютерные игры как одна команда.

М. Ф.: А когда, по вашим прогнозам, ИИ научится рассуждать, думать и действовать как человек?

Дж. Х.: Рассуждения — это одна из тех вещей, с которыми люди действительно хорошо справляются. Пройдет немало времени, прежде чем большие нейронные сети достигнут таких же результатов. И до этого они успеют превзойти людей во многих других вещах.

М. Ф.: Как вы можете охарактеризовать сильный ИИ?

Дж. Х.: Если вы говорите об универсальных роботах, таких как лейтенант-коммандер Дейта, то вряд ли мы будем развивать ИИ

в эту сторону. Кроме того, до создания интеллекта такого уровня еще далеко.

М. Ф.: Возможна ли система, которая сможет пройти тест Тьюринга, длящийся пару часов?

Дж. Х.: Вероятность появления чего-то подобного лет через десять достаточно велика. И возможно, мы до этого не доживем, так как человечество будет уничтожено другими вещами.

М. Ф.: Вы сейчас говорите о таких опасностях, как атомная война или чума?

Дж. Х.: Да, я считаю, что эти вещи угрожают существованию человечества намного сильнее, чем ИИ. Только представьте, что началась мировая война или недовольный аспирант в лаборатории молекулярной биологии создал чрезвычайно заразный смертельный вирус с длительным инкубационным периодом. Я думаю, что беспокоить должно именно это, а не сверхинтеллектуальные системы.

М. Ф.: Но, например, Демис Хассабис из компании DeepMind верит в возможность создания систем, которые вы считаете неосуществимыми.

Дж. Х.: У нас с Демисом разные взгляды на то, каким может быть будущее.

М. Ф.: Я бы все-таки хотел поговорить о связанных с ИИ опасностях. Например, о потенциальном влиянии на рынок труда и экономику. Нужно ли волноваться о наступлении новой промышленной революции?

Дж. Х.: Значительное повышение производительности — это прекрасно. Но его влияние на жизнь людей зависит от социальной системы. Почему проблемой считается именно технический прогресс? Проблема в том, сможем ли мы справедливо распределить блага.

М. Ф.: Но ведь проблема в данном случае связана с исчезновением рабочих мест по причине автоматизации. Может ли ее решить такая вещь, как базовый доход?

Дж. Х.: Да, эта идея кажется мне разумной.

М. Ф.: Но нужны ли для этого политические решения? Некоторые говорят, что можно пустить процесс на самотек и время само расставит все по местам.

Дж. Х.: Я переехал в Канаду, потому что здесь выше ставка налогообложения. С моей точки зрения справедливые налоги — это правильно. Правительство должно помогать всем, в то время как люди имеют право действовать исключительно в собственных интересах. Высокие налоги — один из механизмов, позволяющий реализовывать такую помощь. Кто-то богатеет, а остальные получают пособие из уплаченных им налогов. Но чтобы ИИ приносил пользу всем, предстоит проделать большую работу.

М. Ф.: А что вы скажете о применении ИИ в военных целях?

Дж. Х.: Я думаю, пришло время активности. Нужно сделать так, чтобы международное сообщество относилось к автономному оружию так же, как к химической войне и оружию массового уничтожения.

М. Ф.: Нужен ли мораторий на исследования и разработку оружия с ИИ?

Дж. Х.: Вы не получите моратория на исследования такого типа, точно так же как не было моратория на разработку нервно-паралитических отравляющих веществ. Правда, есть международные механизмы, которые помешали их широкому использованию.

М. Ф.: А что вы думаете о проблемах, связанных с приватностью и прозрачностью?

Дж. Х.: Манипуляция избирателями ради победы на выборах не может не вызывать беспокойства. Создатель компании Cambridge Analytica Боб Мерсер занимался машинным обучением. Компания уже принесла вред, об этом нельзя забывать.

М. Ф.: Тут требуется регулирование?

Дж. Х.: Да, и очень строгое. Это вообще крайне важная тема, но я не эксперт в этих вопросах.

М. Ф.: А глобальная гонка вооружений? Важно ли, чтобы ни одна из стран не вырывалась вперед?

Дж. Х.: Долгое время на мировой арене доминировала Великобритания, которая вела себя не очень гуманно. Потом вперед вышла Америка, которая тоже вела себя примерно так же. Сейчас лидерство может захватить Китай, и вряд ли стоит ожидать от них человеколюбия.

М. Ф.: Нужно ли западным правительствам сделать разработки ИИ национальным приоритетом?

Дж. Х.: Технологии развиваются быстро, и глупо не пытаться идти в ногу со временем. Поэтому вкладывать в ИИ большие средства нужно и важно. Это обыкновенный здравый смысл.

М. Ф.: То есть вы оптимистично смотрите в будущее и считаете, что выгоды от ИИ перевесят потенциальные риски?

Дж. Х.: Я даже не знаю, есть ли вообще какие-то риски. Пока что перед нами стоят проблемы, связанные с социальными системами, а не с технологиями.

М. Ф.: Сейчас в сфере ИИ дефицит кадров, поэтому там рады любым специалистам. Какой совет вы можете дать студентам, которые хотят делать карьеру в этой области?

Дж. Х.: Меня беспокоит нехватка людей, которые критически относятся к основам. Идея капсул была призвана показать, что, возможно, часть общепринятых способов не дает нужных результатов и нужно мыслить шире. Поэтому тем, кто чувствует, что какие-то вещи делаются неправильно, я советую следовать своей интуиции. Студентов и аспирантов я считаю наиболее плодотворным источником новых идей. Они свободно выдвигают действительно новаторские предложения и все время учатся. Те, кто пришел в отрасль, уже получив степень магистра, как правило, не собираются ничего предлагать.

М. Ф.: Создается впечатление, что основные исследования глубокого обучения производятся в Канаде. Это случайность или этому способствуют какие-то факторы?

Дж. Х.: Большая удача, что в Канаде работают Ян Лекун и Иошуа Бенджио. Мы плодотворно сотрудничаем, при поддержке CIFAR, который финансирует фундаментальные исследования, невзирая на высокие риски.

М. Ф.: То есть глубокому обучению удалось остаться на плаву благодаря стратегическим инвестициям канадского правительства?

Дж. Х.: Да. По сути, канадское правительство вкладывает в глубокое обучение по полмиллиона долларов в год, что можно считать разумной инвестицией в индустрию, которая в будущем может принести миллиарды.

М. Ф.: Общаетесь ли вы с вашим коллегой Джорданом Питерсоном? Иногда кажется, что Университет Торонто — это источник дестабилизации…

Дж. Х.: С моей точки зрения, он из тех, кто не понимает, когда лучше промолчать.

> *Дело не в том, что ИИ возненавидит нас за то, что мы поработили его, и не в том, что у него внезапно возникнет искра сознания, которая заставит его взбунтоваться. Просто этот интеллект, скорее всего, будет компетентно преследовать собственные, не совпадающие с нашими цели. И мы окажемся в будущем, сформированном в соответствии с чужими критериями.*

Ник Бостром

ПРОФЕССОР, ОСНОВАТЕЛЬ И ДИРЕКТОР FHI

Ника Бострома знают как одного из ведущих мировых экспертов в области суперинтеллекта и опасностей, которые потенциально несут человечеству ИИ и машинное обучение. Ник — автор бестселлера по версии The New York Times 2014 г. «Искусственный интеллект: этапы, угрозы, стратегии» и около 200 публикаций, в том числе Anthropic Bias («Антропный уклон», 2002 г.), Global Catastrophic Risks («Глобальные риски», 2008 г.), Human Enhancement («Усовершенствование человека», 2009 г.).

Учился в Гетеборгском и Стокгольмском университетах, а также в Королевском колледже Лондона. Степень доктора философии получил в Лондонской школе экономики.

Имеет опыт работы в области физики, ИИ, математической логики и философии. Стал лауреатом премии Eugene R. Gannon Award, которая каждый год вручается одному ученому в мире за достижения в области философии, математики и естественных наук. Его имя находится в списках 100 лучших мыслителей внешней политики и мировых мыслителей журнала Prospect. Самый молодой человек в топ-15 из всех областей и самый высокопоставленный аналитический философ. Его произведения переведены на 24 языка. Сделано более 100 переводов и переизданий его работ.

Мартин Форд: Вы писали об опасностях, связанных с суперинтеллектом — системой, которая может появиться, если сильный ИИ направит свою энергию на самосовершенствование.

Ник Бостром: Да, я описал один из сценариев и связанную с ним проблему, но наступление эры машинного интеллекта может пройти и по другим сценариям, с которыми связаны совсем другие проблемы.

М. Ф.: Вы заострили внимание на проблеме контроля, или выравнивания, то есть на ситуации, когда цели машинного интеллекта могут негативно повлиять на человечество. Расскажите об этом более подробно.

Н. Б.: Что ж, проблема очень продвинутых систем ИИ, которая отличается от других технологий, заключается в том, что не только люди могут злоупотреблять этой технологией. Технология может злоупотреблять собой сама. Другими словами, вы создаете искусственного агента или процесс, который имеет свои собственные цели и задачи, и этот агент имеет все возможности достичь этих целей, потому что в этом сценарии он является сверхразумным. Цели, которые эта мощная система пытается оптимизировать, отличаются от наших человеческих ценностей и, возможно, даже

не соответствуют целям, которых мы хотим достичь в мире. Когда у вас есть люди, пытающиеся достичь чего-то одного, и суперинтеллектуальная система, пытающаяся достичь чего-то другого, вполне возможно, что суперинтеллект будет побеждать и добиваться своего.

Дело не в том, что ИИ возненавидит нас за то, что мы поработили его, и не в том, что у него внезапно возникнет искра сознания, которая заставит его взбунтоваться. Просто этот интеллект, скорее всего, будет компетентно преследовать собственные, не совпадающие с нашими цели. И мы окажемся в будущем, сформированном в соответствии с чужими критериями. Соответственно, чтобы решить проблему контроля, или выравнивания, нужно найти способы проектировать системы ИИ как продолжения человеческой воли. Чтобы поведение этих систем формировалось человеческими намерениями, а не какими-то случайными результатами внутри самих систем.

М. Ф.: В своих интервью вы приводили в пример систему, производящую скрепки, которая достигает поставленной перед ней цели с помощью сверхинтеллектуальных способностей, но в конечном итоге превращает в скрепки всю Вселенную, так как лишена здравого смысла. Это иллюстрация проблемы выравнивания?

Н. Б.: Пример скрепки иллюстрирует широкую категорию возможных сбоев, когда вы просите систему сделать что-то одно, и, возможно, вначале все получается довольно хорошо, но затем она приходит к выводу, который находится вне нашего контроля. Пример, в котором вы разрабатываете ИИ для управления фабрикой скрепок, карикатурен, но хорошо передает суть. Сначала этот ИИ глуп, но чем умнее он становится, тем лучше работает. Владелец фабрики очень доволен и хочет добиться большего прогресса. Однако когда ИИ становится достаточно умным, он осознает, что существуют другие способы создания еще большего количества скрепок в мире, например, он может забрать контроль у людей

и превратить всю планету в скрепки или космические зонды, которые превратят всю вселенную в скрепки.

Суть в том, что при постановке цели, включающей увеличение количества, нужно быть крайне осторожными в формулировках.

М. Ф.: Почему основные обсуждения посвящены способам достижения цели? Я не слышал примера, в котором система просто поменяла бы свою цель. Люди поступают так сплошь и рядом!

Н. Б.: Хотя суперинтеллект обладает способностью менять свои цели, следует учитывать, что он делает выбор между новыми и текущими целями. В большинстве ситуаций перемена цели кажется ИИ плохим стратегическим шагом — он понимает, что в результате не окажется агента, преследующего его текущую цель, и отдает ей приоритет. Такая достаточно сложная система рассуждений позволяет достичь стабильности внутренней цели.

Мы, люди, противоречивы по своей природе. Кажется, что иногда мы решаем изменить наши цели. Но о решении на самом деле речи не идет — цели меняются иначе. Более того, под целями люди подразумевают не фундаментальные критерии оценки вещей, а желание достичь какого-то результата, которое может пройти при изменении обстоятельств.

М. Ф.: Но ведь многие исследования в области ИИ базируются на нейробиологии, а в машинный интеллект мы внедряем свои, человеческие идеи. Представьте систему, имеющую в распоряжении все человеческие знания. В мозге человека могут возникать разные патологии, существуют даже препараты, влияющие на работу мозга. Откуда мы знаем, что у машин не может быть ничего похожего?

Н. Б.: Я допускаю подобную возможность, особенно на ранних этапах разработки ИИ, когда машина еще не поняла, как ме-

нять себя, не нанося при этом себе вреда. Нужно разработать технологию, позволяющую предотвращать изменения в целеполагании. Я ожидаю, что достаточно развитая система сможет разработать технологию для обеспечения собственной целостности и даже сделать этот подход приоритетным. Но пока система недостаточно развита, существует вероятность ее самоповреждения.

М. Ф.: Когда речь заходит о машинах, делающих не то, что мы хотим, меня смущает слово «мы». Ведь не существует универсального набора человеческих желаний и ценностей.

Н. Б.: Большие задачи принято решать, разбивая их на более мелкие. Ставя вопрос, как добиться согласованности ИИ с любыми человеческими ценностями, чтобы заставить его действовать в соответствии с пожеланиями разработчиков, вы выделяете одну подзадачу. Без ее решения нет смысла разбирать политические проблемы. Надо сначала заняться технологией, а потом уже спорить о ее применении.

М. Ф.: Я правильно понимаю, что сейчас в FHI и в других исследовательских компаниях, таких как OpenAI, решается проблема технического контроля, то есть ищется ответ на вопрос, как построить машину, работающую в строгом соответствии с поставленной целью.

Н. Б.: Именно так. У нас над этим работает отдельная команда. Есть и команда, которая занимается проблемами управления, связанными с достижениями в области ИИ.

М. Ф.: Хватит ли у вас ресурсов для управления ИИ или нужна поддержка правительства?

Н. Б.: Мне бы хотелось, чтобы больше ресурсов выделялось на решение вопросов безопасности ИИ. Этим занимаемся не только мы, например, в компании DeepMind есть отдельная группа. Вообще

количество талантов и денег в сфере ИИ растет, хотя в абсолютном выражении эта область все еще очень неразвита.

М. Ф.: Нужно ли к проблемам, которые порождает суперинтеллект, привлекать внимание общества?

Н. Б.: Пока в этом нет смысла, потому что не совсем ясно, какая именно помощь будет полезной в данный момент. Сейчас я не вижу необходимости в каких-либо правилах, касающихся машинного суперинтеллекта. Регулирование понадобится для приложений на базе ИИ, которые появятся в ближайшем будущем. Мне кажется, суперинтеллект не входит в компетенцию политиков, так как их в основном заботит то, что может произойти во время их пребывания в должности.

М. Ф.: Могут ли усугубить ситуацию выступления Илона Маска, который утверждает, что суперинтеллект опаснее, чем Северная Корея?

Н. Б.: Преждевременные выступления такого рода могут привести к тому, что голоса выступающих за осторожность и глобальное сотрудничество отойдут на второй план. И это действительно может ухудшить ситуацию. Мне кажется, обращать внимание правительства на проблемы суперинтеллекта нужно тогда, когда возникнет необходимость. Но до этого еще предстоит проделать огромную работу.

М. Ф.: Как развивалась ваша карьера в сфере ИИ?

Н. Б.: Тема ИИ интересовала меня, сколько я себя помню. В университете я изучал ИИ, а затем вычислительную нейробиологию и теоретическую физику. Во-первых, я считал, что технология ИИ в конечном итоге изменит мир, а во-вторых, я хотел понять, каким образом возникает мышление.

Первые работы на тему суперинтеллекта я опубликовал в середине 1990-х гг. А в 2006 г. появилась возможность открыть FHI

в Оксфордском университете и исследовать влияние новых технологий на наше будущее. Особое внимание мы уделяем машинному интеллекту. В 2014 г. по результатам работы появилась книга «Искусственный интеллект: этапы, угрозы, стратегии». В настоящее время в институте две рабочие группы. Одна фокусируется на решении проблемы выравнивания, разрабатывая алгоритмы для масштабируемых методов управления. Вторая занимается вопросами управления, политики, этики и социальных последствий, которые несут в себе достижения в области машинного интеллекта.

М. Ф.: При этом в своей работе вы рассматриваете опасности, связанные не только с ИИ?

Н. Б.: Да. А также мы исследуем открывающиеся возможности и технологические преимущества.

М. Ф.: Какие еще опасности вы рассматриваете и почему концентрируетесь на машинном интеллекте?

Н. Б.: В FHI мы изучаем масштабные вопросы, касающиеся судьбы разумной жизни на Земле. Мы рассматриваем экзистенциальные угрозы и то, что может навсегда сформировать наше развитие в будущем. Технологии я считаю наиболее вероятным источником таких фундаментальных преобразований. Разумеется, далеко не со всеми технологиями связаны экзистенциальные опасности и возможности, как с ИИ. Мы изучаем опасности от биотехнологий. Эти исследования необходимо объединить в макростратегию.

Почему мы выделяем ИИ? С моей точки зрения, если ИИ успешно достигнет первоначальной цели, которая сводится не только к автоматизации конкретных задач, но и к копированию универсальной способности к обучению и планированию, это станет последним изобретением, которое сделают люди. Сильно изменятся все технологические и прочие области, в которых полезен человеческий интеллект.

М. Ф.: Входит ли в ваш список экзистенциальных опасностей изменение климата?

Н. Б.: Нет, потому что мы предпочитаем сосредоточиться на областях, в которых могут иметь значение наши усилия. А проблемами, вызванными изменением климата, сейчас занимается множество специалистов. Кроме того, мне трудно представить, что увеличение температуры на несколько градусов может привести к исчезновению человеческого рода. Поэтому влияние климата мы рассматриваем только в рамках общей картины проблем, с которыми может столкнуться человечество.

М. Ф.: Итак, с вашей точки зрения, опасности, связанные с сильным ИИ, выше связанных с изменением климата. Получается, что ресурсы и инвестиции распределяются неоптимальным образом.

Н. Б.: Я действительно думаю, что ресурсы распределяются не совсем правильно, причем не только между этими двумя областями. Мне в принципе кажется, что мы не очень мудро распределяем внимание.

Если оглянуться назад, мы увидим, что в любой момент времени существовала какая-то глобальная проблема, которой были озабочены все образованные люди, а потом все менялось. Сто лет назад все были озабочены проблемой вырождения, и все говорили об ухудшении человеческой породы. Во время холодной войны на первый план вышел ядерный армагеддон, а затем зашла речь о перенаселении. Сейчас все обеспокоены глобальным потеплением, хотя последние пару лет все больше начинают говорить об опасностях, которые несет ИИ.

М. Ф.: Это может быть связано с выступлениями таких людей, как Илон Маск?

Н. Б.: Пока я вижу в этом положительную сторону. Когда я писал книгу, оказалось, что тема ИИ мало кого интересует. Даже среди

специалистов в этой сфере немногие думали о том, что может произойти, если появится ИИ. Серьезно говорить на эту тему было невозможно.

А вследствие того, что к этой теме было привлечено внимание, теперь можно проводить исследования и публиковать статьи по таким вопросам, как проблема выравнивания. Не думаю, что столько талантливых людей пришло бы в эту область, не начнись громкое обсуждение. Теперь же задача состоит в том, чтобы направить появившуюся озабоченность и интерес в конструктивное русло и продолжить работу.

М. Ф.: Верно ли я понимаю, что вас беспокоят опасности, связанные с появлением сильного ИИ? Обычный же ИИ, несмотря на свою значительность, не угрожает существованию людей.

Н. Б.: Именно так. Разумеется, нас интересуют и приложения на базе ИИ, которые могут появиться в ближайшем будущем. Мне кажется, проблема возникает, когда начинают путать краткосрочный и долгосрочный контексты.

М. Ф.: А о чем нужно беспокоиться в ближайшие пять лет?

Н. Б.: На мой взгляд, выгоды от вещей, которые появятся в ближайшее время, перевешивают недостатки. Достаточно посмотреть на экономику и на все те области, где могут пригодиться более умные алгоритмы. Даже простой алгоритм прогнозирования кривых спроса, работающий в фоновом режиме в большом логистическом центре, позволит сократить запасы товаров и снизить цены.

Нейронные сети могут распознавать опухоли на рентгеновских снимках и помогать в постановке диагнозов. Работая в фоновом режиме, они позволят оптимизировать обслуживание пациентов. Предприниматели найдут массу возможностей. А с научной точки зрения интересно хоть немного понять, как работает интеллект и каким образом осуществляется восприятие.

М. Ф.: Многих беспокоит такая вещь, как оружие, самостоятельно выбирающее цель. Поддерживаете ли вы запрет на оружие такого типа?

Н. Б.: Хотелось бы, чтобы мир смог избежать новой гонки вооружений с тратой огромных сумм на совершенствование роботов-убийц. Вообще говоря, я бы предпочел, чтобы машинный интеллект использовался исключительно в мирных целях. Но если смотреть более масштабно, сложно понять, что именно нужно запрещать.

Есть мнение, что следует запретить автономные дроны, самостоятельно выбирающие цель. Но в качестве альтернативы у нас есть девятнадцатилетний парень, которому поручено нажимать кнопку, когда на экране появляется надпись «Огонь». Я не понимаю, чем это отличается от полностью автономной системы. Куда важнее, чтобы был человек, который будет отвечать за случившееся, если что-то пойдет не так.

М. Ф.: Но можно представить ситуации, в которых автономная машина предпочтительнее. Например, в США отмечались случаи расизма со стороны полицейских. В правильно спроектированной роботизированной системе предвзятость можно исключить. Кроме того, робота-полицейского можно запрограммировать так, чтобы он начинал стрелять только после того, как в него попадет пуля.

Н. Б.: Мне бы хотелось, чтобы люди вообще перестали воевать. Но раз войны неизбежны, наверное, лучше, когда машины убивают другие машины. Если речь идет об ударах по конкретным целям, важна возможность наносить точные удары, позволяющие избежать сопутствующего ущерба. Вот почему я говорю, что, если принять во внимание специфику, общие расчеты сильно усложняются, и непонятно, как должно выглядеть соглашение в отношении автономного оружия.

Этические вопросы возникают и в других областях применения, таких как наблюдение, управление потоками данных, маркетинг и реклама. Все это в долгосрочной перспективе тоже может повлиять на человеческую цивилизацию.

М. Ф.: То есть нужны законы, регулирующие применение этих технологий?

Н. Б.: Без регулирования действительно не обойтись. Вы же не хотите, чтобы с дрона, оснащенного программой распознавания лиц, можно было с большой дистанции убить человека. Точно так же вы не хотите, чтобы дроны, летающие через аэропорт, задерживали самолеты. Я уверен, что по мере увеличения количества беспилотников потребуется военизированная структура для их контроля.

М. Ф.: С момента публикации вашей книги прошло около пяти лет. Ваши ожидания оправдались?

Н. Б.: Развитие происходит быстрее, чем я предсказывал.

М. Ф.: В книге вы писали, что компьютер победит чемпиона мира по игре го где-то через десять лет, примерно в 2024 г. Но это случилось через два года.

Н. Б.: Я писал, исходя из предположения, что развитие будет иметь стабильную скорость. Но прогресс пошел намного быстрее, отчасти благодаря особым усилиям, которые фирма DeepMind приложила к решению этой конкретной задачи. Были привлечены хорошие специалисты и выделены вычислительные мощности. В результате нам продемонстрировали впечатляющие возможности систем глубокого обучения.

М. Ф.: А какие препятствия стоят на пути к сильному ИИ?

Н. Б.: Сильный ИИ требует более совершенных методов обучения без учителя. Ведь только небольшая часть знаний и навыков

взрослого человека формируется при помощи явных инструкций. В основном же мы просто наблюдаем за происходящим вокруг, используя для улучшения модели мира сенсорный канал. В детстве мы много пробуем и ошибаемся — и таким способом учимся.

Для высокоэффективных интеллектуальных систем нужны алгоритмы, умеющие использовать немаркированные данные. Большую часть своих знаний о мире люди склонны систематизировать в виде причинно-следственных связей, а в современных нейронных сетях этого практически нет. В основном достигается нахождение статистических закономерностей в сложных моделях, а не превращение этих моделей в объекты, влияющие друг на друга.

Кроме того, нужен прогресс в планировании и ряде других областей. Доступные методы не очень хорошо справляются с существующими задачами на пути к пониманию человеческого интеллекта.

М. Ф.: Одна из немногих компаний, специализирующихся на разработке сильного ИИ, — это DeepMind. Есть ли другие игроки, которые, на ваш взгляд, могут с ней конкурировать?

Н. Б.: Безусловно, компания DeepMind — один из лидеров, но в рамках проекта Google Brain тоже работает исследовательская группа мирового класса. Собственные ИИ-лаборатории есть и у таких крупных технологических компаний, как Facebook, Baidu и Microsoft.

Занимаются этим и в академических кругах. Университеты в Монреале и Торонто одни из лучших в мире в сфере глубокого обучения. Много специалистов работает в таких университетах, как Беркли, Оксфорд, Стэнфорд и Карнеги — Меллона. Западными странами дело не ограничивается. Китай также вкладывает значительные средства в наращивание внутреннего потенциала.

М. Ф.: Но они не специализируются конкретно на сильном ИИ.

Н. Б.: Четкую границу в данном случае провести сложно. Но если брать группы, занятые именно разработкой сильного ИИ, то кроме DeepMind можно упомянуть OpenAI.

М. Ф.: Тест Тьюринга — это единственный способ определить, получилось ли создать сильный ИИ?

Н. Б.: Полномасштабная версия теста вполне подходит для подтверждения завершенности ИИ-полной задачи. Если же хочется оценить скорость движения к цели и понять, в каком направлении работать дальше, тест Тьюринга не поможет.

М. Ф.: Потому что становится субъективным?

Н. Б.: Да. Конечно, тест можно применить корректно даже в таких вопросах, но это слишком сложно, и мы пока не знаем, как это реализовать. Когда мы пытаемся оценить локальный прогресс системы, на результаты теста Тьюринга начинает влиять множество стандартных ответов, а также другие уловки и приемы. Этот тест не сможет показать, продвинулись ли ваши разработки в сторону сильного ИИ. Для этого нужны другие критерии, которые будут учитывать движение в нужную сторону.

М. Ф.: У интеллектуальной системы может автоматически появиться сознание?

Н. Б.: Что вы подразумеваете под сознанием? Например, этим термином называют функциональную форму самоощущения, то есть возможность смоделировать себя в мире и оценить влияние окружающей среды и времени.

Еще под сознанием подразумевают результат накопления эмпирического опыта, который, как принято считать, имеет значение для морали. Например, аморально причинять страдания живым существам, причем страдание — это субъективное чувство.

Неизвестно, возникнет ли такой феноменальный опыт автоматически как следствие повышения эффективности системы. Можно даже создать механические системы, лишенные сенсорных, чувственных явлений любого рода, то есть того, что философы называют термином «квалиа». Мы не знаем, какие вещи необходимы и достаточны для морально значимых форм сознания, поэтому нужно принять как аксиому, что машинный интеллект может получить сознание задолго до достижения им человеческого уровня.

Считается, что соответствующий опыт есть не только у людей. Поэтому существуют протоколы и руководства для медицинских опытов с мышами. Нужно помнить, что машина может достигнуть уровня сознания, обеспечивающего некий моральный статус, что ограничит наши действия по отношению к ней.

М. Ф.: То есть рискуют в данном случае обе стороны? Нас беспокоит вред, который может причинить ИИ, но при этом мы сами можем поработить сущность, обладающую сознанием, или заставить ее страдать. Мне кажется, узнать, обладает ли машина сознанием, невозможно. Я считаю, что у вас есть сознание, потому что вы одного со мной вида, а в наличии у себя сознания я уверен. Но у меня нет такой же связи с машиной.

Н. Б.: Да, это действительно крайне сложный вопрос. Ведь даже принадлежность к человеческому виду не говорит о наличии сознания. Человек может находиться в коме, на стадии зародыша, под наркозом. При этом многие уверены в наличии сознательного опыта у определенных животных. То есть мы проецируем идею сознания за пределы нашего собственного вида, но мне кажется, что на цифровой интеллект, если таковой появится, эти представления, как и моральные нормы, вряд ли распространятся.

Даже с животными мы обращаемся плохо. Особенно это касается производства мяса, а ведь животные могут выражать эмоции

и кричать! Кто заставит людей осознать, что внутри микропроцессора, где происходят невидимые реакции, может быть разум, с которым нужно считаться? Особенно на фоне таких проблем, как алгоритмическая дискриминация или дроны-убийцы. Поэтому первым делом нужно сделать так, чтобы эта тема перестала относиться к абстракциям, о которых разговаривают только профессиональные философы.

М. Ф.: А как с вашей точки зрения ИИ может повлиять на рынок труда и экономику?

Н. Б.: В ближайшей перспективе сильных воздействий на рынок труда не предвидится, так как для масштабного развертывания автоматизированных систем требуется время. Но постепенно достижения в области машинного обучения начнут все сильнее влиять на эту сферу, ведь ИИ способен взять на себя практически все задачи. В некоторых отношениях конечная цель состоит в освобождении людей от необходимости работать. Ведь мы развиваем технологии и внедряем автоматизацию, чтобы получать определенный результат с меньшими усилиями.

М. Ф.: Но это же утопия. Как я понимаю, вы поддерживаете идею базового дохода, который гарантирует всем и каждому возможность пользоваться плодами прогресса?

Н. Б.: Да, со временем желательно прийти к чему-нибудь в этом роде. Если действительно получится создать ИИ, и мы решим проблему технического контроля, произойдет взрывной экономический рост. Этого хватит, чтобы обеспечить всем достойную жизнь. Развивая суперинтеллект, мы все рискуем, нравится нам это или нет. Поэтому, если дела пойдут хорошо, каждый должен получить некоторую выгоду.

Я думаю, что именно так должен использоваться машинный суперинтеллект. Если все получится, выигрыш будет настолько большим, что мы просто обязаны гарантировать всем фантастиче-

ское качество жизни: не только экономическую выгоду (в форме универсального базового дохода или по какой-то другой схеме), но совершенные технологии, лучшее здравоохранение и т. п.

М. Ф.: А если сильный ИИ первым создаст Китай? Имеют ли значение культурные ценности страны, в которой развивается технология?

Н. Б.: Я думаю, не очень важно, в какой культуре сильный ИИ появится первым. Куда важнее вопрос, насколько компетентны люди или группы, которые его развивают, и имеют ли они возможность быть осторожными. Это проблема любой гонки. Стремясь к результату, люди пренебрегают правилами безопасности. И побеждает тот, кто тратит на безопасность меньше всего усилий.

Хотелось бы, чтобы тот, кто первым разработает суперинтеллект, имел возможность в конце сделать паузу месяцев на шесть, а лучше на пару лет, чтобы перепроверить свои системы и установить любые дополнительные средства защиты, которые сможет придумать. И только после этого медленно и осторожно стал расширять возможности системы до сверхчеловеческого уровня. Поэтому важно максимально смягчать конкуренцию.

М. Ф.: Если интеллект сможет рекурсивно улучшать сам себя, первопроходец получает огромное преимущество. Соответственно, существует огромный стимул именно для гонки.

Н. Б.: Да, это так. Но я думаю, что тема гарантированного достижения глобального блага может снизить интенсивность гонки. Нужно, чтобы все участники осознавали, что проигравших не будет.

М. Ф.: Это потребует международного сотрудничества, в котором человечество пока не очень преуспело. Если сравнить с запретом на химическое оружие и с актом о нераспространении ядерного оружия, получится, что в случае с ИИ проверить соблюдение соглашения будет еще сложнее.

Н. Б.: В чем-то это окажется сложнее, в чем-то проще. Люди играют в эти игры из-за ограниченных ресурсов. У кого-то ресурсы есть, а у кого-то еще нет. ИИ может привести к изобилию во многих отношениях, что позволит облегчить установление договоренностей.

М. Ф.: Думаете, мы решим эти проблемы?

Н. Б.: Внутри у меня надежда и страх. Но хотелось бы подчеркнуть именно положительные стороны, как в краткосрочной, так и в долгосрочной перспективе. Из-за моей работы и книги меня всегда спрашивают о рисках и недостатках технологии, но я надеюсь, что она сможет стать благом для всего мира.

> *Человек может научиться водить автомобиль за 15 часов тренировок, ни во что не врезавшись. Если использовать существующие методы обучения с подкреплением, машине, чтобы научиться ездить без водителя, придется 10 тысяч раз упасть с обрыва, прежде чем она поймет, как этого избежать.*

Ян Лекун

ВИЦЕ-ПРЕЗИДЕНТ И ОСНОВАТЕЛЬ ЛАБОРАТОРИИ ИССЛЕДОВАНИЯ ИИ В FACEBOOK (FAIR), ПРОФЕССОР COMPUTER SCIENCE В НЬЮ-ЙОРКСКОМ УНИВЕРСИТЕТЕ

Вместе с Джеффри Хинтоном и Иошуа Бенджио Ян Лекун входит в группу исследователей, усилия и настойчивость которых привели к нынешней революции в отношении к нейронным сетям и глубокому обучению. Работая в Лабораториях Белла, он изобрел сверточные нейронные сети. Диплом инженера-электрика получил в Париже в ESIEE, а докторскую степень в области computer science — в Университете Пьера и Марии Кюри. После аспирантуры работал в Лаборатории Джеффри Хинтона в Университете Торонто.

Мартин Форд: Взрыв интереса к глубокому обучению последние 10 лет — это следствие одновременного совершенствования нейронных сетей, увеличения мощности компьютеров и количества доступных данных?

Ян Лекун: Да, но процесс был более обдуманным. Появившийся в 1986–87 гг. алгоритм обратного распространения дал возможность обучать многослойные нейронные сети. Это вызвало волну интереса, которая продержалась вплоть до 1995 г. В 2003 г. Джеффри Хинтон, Иошуа Бенджио и я придумали план, как возобновить интерес сообщества к этим методам, потому что были уверены в их неминуемой победе. Так что можно сказать, что имел место умышленный сговор.

М. Ф.: Вы уже тогда понимали все перспективы? Сейчас ИИ и глубокое обучение считают синонимами.

Я. Л.: И да, и нет. Мы знали, что методы лягут в основу компьютерного зрения, распознавания речи и, возможно, пары других вещей, но никто не ожидал, что они распространятся на понимание естественного языка, робототехнику, анализ медицинской визуализации и даже поспособствуют появлению беспилотных автомобилей. В начале 1990-х гг. я думал, что движение к этим вещам будет более плавным, а появятся они немного раньше. Нас же ждала революция, случившаяся примерно в 2013 г.

М. Ф.: А как возник ваш интерес к ИИ и машинному обучению?

Я. Л.: Я с детства интересовался наукой, техникой и глобальными вопросами о зарождении жизни, интеллекта, происхождении человечества. Идея ИИ привела меня в восторг. Но в 1960–70 х гг. во Франции этим никто не занимался, поэтому после школы я пошел учиться на инженера.

В 1980 г. мне очень понравилась книга по философии *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*

(«Язык и обучение: дискуссия между Жаном Пиаже и Ноамом Хомским»), в которой создатель теории когнитивного развития и лингвист обсуждали природу и воспитание, а также зарождение языка и интеллекта.

На стороне Пиаже выступал профессор MIT Сеймур Пейперт, который стоял у истоков машинного обучения и в конце 1960-х гг. фактически способствовал прекращению работ с нейронными сетями. И вот спустя 10 лет он превозносил так называемый персептрон — очень простую модель машинного обучения, которая появилась в 1950-х гг. и над которой он работал в 1960-х гг. Так я впервые познакомился с концепцией обучения машин и был ею абсолютно очарован. Способность к обучению я считал неотъемлемой частью интеллекта.

Студентом я прочитал по машинному обучению все, что удалось найти, и сделал несколько проектов по этой теме. Оказалось, на Западе никто не работает с нейронными сетями. Над тем, что позже стало называться этим термином, трудились несколько японских исследователей. У нас же эта тема никого не интересовала, отчасти из-за вышедшей в конце 1960-х гг. книги Пейперта и Минского.

Я начал самостоятельные исследования и в 1987 г. защитил докторскую диссертацию *Modeles connexionnistes de l'apprentissage* («Коннекционистские модели обучения»). Мой руководитель Морис Милгрэм этой темой не занимался и прямо сказал мне, что может официально стать моим консультантом, но ничем не сможет помочь технически.

В начале 1980-х гг. я обнаружил сообщество людей, которые работали над нейронными сетями, и связался с ними. В итоге параллельно Дэвиду Румельхарту и Джеффри Хинтону я открыл такую вещь, как метод обратного распространения ошибки.

М. Ф.: То есть в начале 1980-х гг. в Канаде велись многочисленные исследования в этой области?

Я. Л.: Нет, все происходило в США. В Канаде такие исследования тогда еще не велись. В начале 1980-х гг. Джеффри Хинтон был сотрудником Калифорнийского университета в Сан-Диего, где работал с такими специалистами по когнитивной психологии, как Дэвид Румельхарт и Джеймс Макклелланд. В результате появилась книга, объясняющая психологию при помощи простых нейронных сетей и компьютерных моделей. Затем Джеффри стал доцентом в Университете Карнеги — Меллона. В Торонто он переехал только в 1987 г. Тогда же в Торонто перебрался и я, и в течение года работал в его лаборатории.

М. Ф.: В начале 1980-х гг. я был студентом, изучавшим вычислительную технику, и не помню, чтобы где-то применялись нейронные сети. Сейчас ситуация резко изменилась.

Я. Л.: Нейронные сети не просто оказались на обочине науки. В 1970-х гг. и начале 1980-х гг. их фактически предали анафеме. Статьи отклонялись за одно упоминание нейронных сетей.

Известна статья *Optimal Perceptual Inference* («Оптимальный перцептивный вывод»), которую в 1983 г. опубликовали Джеффри Хинтон и Терри Сейновски. Чтобы описать в ней одну из первых моделей глубокого обучения и нейронной сети, они использовали кодовые слова, даже в названии.

М. Ф.: Вы известны как автор сверточной нейронной сети. Объясните, пожалуйста, что это такое?

Я. Л.: Изначально эта нейронная сеть была оптимизирована под распознавание объектов на изображениях. Но оказалось, что ее можно применить к широкому кругу задач, например распознаванию речи и машинному переводу. Идеей для ее создания послужили особенности зрительной коры мозга животных и людей, изученные в 1950–60-х гг. Дэвидом Хьюбелом и Торстеном Визелом, позднее получившими Нобелевскую премию в области нейробиологии.

Сверточная сеть — это особый способ соединения нейронов, которые не являются точной копией биологических нейронов. В первом слое — слое свертки — каждый нейрон связан с небольшим количеством пикселов изображения и вычисляет взвешенную сумму своих входных данных. В процессе обучения веса меняются. Группы нейронов видят небольшие участки изображения. Если нейрон обнаруживает определенный признак на одном участке, другой нейрон обнаружит точно такой же признак на соседнем участке, а все остальные нейроны — в остальных участках изображения. Математическая операция, которую нейроны выполняют вместе, называется дискретной сверткой. Отсюда название.

Затем идет нелинейный слой, где каждый нейрон включается или выключается, в зависимости от того, выше или ниже заданного порога оказалась вычисляемая слоем свертки взвешенная сумма. Наконец, третий слой выполняет операцию субдискретизации, чтобы убедиться, что небольшое смещение или деформация входного изображения не сильно меняет результат на выходе. Это обеспечивает независимость от деформаций входного изображения.

По сути, сверточная сеть — это стек, организованный из слоев свертки, нелинейности и субдискретизации. Когда они сложены, появляются нейроны, распознающие объекты. Например, нейрон, который включается при нахождении лошади на изображении, другой нейрон — для автомобилей, третий — для людей и так далее, для всех нужных вам категорий.

При этом то, что делает нейронная сеть, определяется силой связей между нейронами, то есть весами. И эти веса не запрограммированы, а являются результатом обучения.

Сети показывается изображение лошади, и, если она не отвечает «лошадь», ее информируют, что это неправильно, и подсказывают правильный ответ. После этого с помощью алгоритма обратного

распространения ошибки сеть корректирует веса всех соединений, чтобы в следующий раз при демонстрации такого же изображения результат был ближе к нужному. При этом приходится демонстрировать ей тысячи изображений.

М. Ф.: Это обучение с учителем? Как я понимаю, сейчас это доминирующий подход.

Я. Л.: Именно так. Почти все современные приложения глубокого обучения используют обучение с учителем. Магия в том, что обученная сеть по большей части дает правильные ответы даже для изображений, которых ей раньше не показывали. Но нуждается в огромном количестве примеров.

М. Ф.: А чего можно ожидать в будущем? Можно ли будет учить машину как ребенка, которому достаточно один раз показать кошку и назвать ее?

Я. Л.: На самом деле вы не совсем правы. Первые тренировки сверточной сети действительно проходят на миллионах изображений различных категорий. А потом, если нужно добавить новую категорию, например научить компьютер распознавать кошек, для этого достаточно нескольких образцов. Ведь сеть уже обучена распознавать объекты практически любого типа. Дополнения к обучению касаются пары верхних слоев.

М. Ф.: Это уже похоже на то, как учатся дети.

Я. Л.: Нет, к сожалению, это совсем не похоже. Дети получают большую часть информации до того, как кто-то скажет им: «Это кошка». В первые несколько месяцев жизни дети учатся, не имея понятия о языке. Они узнают устройство мира, просто наблюдая за миром и немного взаимодействуя с ним. Такой способ накопления знаний машинам недоступен. Как это назвать, непонятно. Некоторые используют провокационный термин «обучение без учителя».

Иногда это называют предвосхищающим, или индуктивным, обучением. Я называю это самообучением. При обучении этого типа не идет речи о подготовке к выполнению какой-то задачи, это просто наблюдение за миром и тем, как он функционирует.

М. Ф.: А обучение с подкреплением в эту категорию попадает?

Я. Л.: Нет, это совсем другая категория. По сути, выделяют три основные категории: обучение с подкреплением, обучение с учителем и самообучение.

Обучение с подкреплением происходит методом проб и ошибок и хорошо работает для игр, где можно делать сколько угодно попыток. Хорошая производительность AlphaGo была достигнута после того, как машина сыграла больше игр, чем все человечество за последние три тысячи лет. К задачам из реального мира такой подход нецелесообразен.

Человек может научиться водить автомобиль за 15 часов тренировок, ни во что не врезавшись. Если использовать существующие методы обучения с подкреплением, машине, чтобы научиться ездить без водителя, придется 10 тысяч раз упасть с обрыва, прежде чем она поймет, как этого избежать.

М. Ф.: Мне кажется, что это аргумент в пользу моделирования.

Я. Л.: Скорее, это подтверждение того, что тип обучения, которым пользуются люди, сильно отличается от обучения с подкреплением. Это похоже на обучение с подкреплением на базе моделей. Ведь человек, садясь за руль впервые, имеет модель мира и может предсказывать последствия своих действий. Как заставить машину самостоятельно изучать прогностические модели — это главная нерешенная проблема.

М. Ф.: Именно с этим связана ваша работа в Facebook?

Я. Л.: Да, это одна из вещей, над которыми мы работаем. Еще мы обучаем машину наблюдать за разными источниками данных. Строим модель мира, надеясь на отражение в ней здравого смысла, чтобы потом использовать ее как прогностическую.

М. Ф.: Некоторые считают, что одного глубокого обучения недостаточно, и в сетях изначально должна быть структура, отвечающая за интеллект. А вы, похоже, убеждены, что интеллект может органически появиться из относительно универсальных нейронных сетей.

Я. Л.: Вы преувеличиваете. С необходимостью структуры согласны все, вопрос в том, как она должна выглядеть. А говоря о людях, которые считают, что должны быть структуры, обеспечивающие логическое мышление и способность к аргументации, вы, вероятно, имеете в виду Гари Маркуса и, возможно, Орена Этциони. С Гари мы спорили на эту тему сегодня утром. Его мнение не очень хорошо воспринимается в сообществе, потому что, не сделав ни малейшего вклада в глубокое обучение, он критически писал о нем. Орен работал в этой сфере некоторое время и при этом высказывается значительно мягче.

Фактически, сама идея сверточных сетей возникла как попытка добавить в нейронные сети структуру. Вопрос в том, какую: позволяющую машине манипулировать символами или, например, соответствующую иерархическим особенностям языка?

Многие мои коллеги, в том числе Джеффри Хинтон и Иошуа Бенджио, согласны с тем, что рано или поздно мы сможем обойтись без структур. Они могут принести пользу в краткосрочной перспективе, потому что пока не придуман способ самообучения. Этот момент можно обойти, привязав все к архитектуре. Но микроструктура коры, как визуальной, так и префронтальной, кажется полностью однородной.

М. Ф.: А мозг использует что-то похожее на метод обратного распространения ошибки?

Я. Л.: Это неизвестно. Может оказаться, что это не обратное распространение в том виде, как мы его знаем, а похожая на него форма аппроксимации оценки градиента. Над биологически правдоподобными формами оценки градиента работал Иошуа Бенджио. Существует вероятность того, что мозг оценивает градиент какой-либо целевой функции.

М. Ф.: Над какими еще важными вещами ведется работа в компании Facebook?

Я. Л.: Мы занимаемся множеством фундаментальных исследований, а также вопросами машинного обучения, поэтому в основном имеем дело с прикладной математикой и оптимизацией. Ведется работа над обучением с подкреплением и над так называемыми порождающими моделями, которые представляют собой форму самообучения или предвосхищающего обучения.

М. Ф.: Разрабатывает ли компания Facebook системы, умеющие поддерживать разговор?

Я. Л.: Фундаментальные темы исследований я перечислил выше, но есть еще и множество областей их применения. Facebook активно ведет разработки в области компьютерного зрения, и можно утверждать, что у нас лучшая в мире исследовательская группа. Мы много работаем и над обработкой текстов на естественном языке. Сюда относится перевод, обобщение, категоризация (выяснение, о какой теме идет речь) и диалоговые системы для виртуальных помощников, систем вопросов и ответов и т. п.

М. Ф.: Как вы думаете, появится ли однажды ИИ, который сможет пройти тест Тьюринга?

Я. Л.: В какой-то момент это случится, но я не считаю тест Тьюринга хорошим критерием: его легко обмануть, и он в некоторой степени устарел. Многие забывают или отказываются верить, что язык — это вторичное явление по отношению к интеллекту.

Посмотрите на орангутанов. Они в изрядной степени обладают здравым смыслом, имеют хорошие модели мира и могут создавать инструменты, как люди. Но при этом у них нет языка. Они не социальные животные и едва взаимодействуют с другими представителями вида, если не брать невербальное общение матери и детеныша. Существует целый пласт интеллекта, не имеющий ничего общего с языком. И сводя проверку ИИ к прохождению теста Тьюринга, мы этот пласт игнорируем.

М. Ф.: Какие препятствия стоят на пути к созданию сильного ИИ?

Я. Л.: Я думаю, что мы пока не видим всего массива проблем, с которым нам предстоит столкнуться в процессе работы. Но первым делом нужно выяснить, каким способом дети и животные в первые дни, недели и месяцы жизни познают устройство мира.

Именно в это время ребенок узнает, что мир трехмерен. Замечает, что при движении головой объекты перемещаются перед ним. Получает представление о постоянстве предметов, когда видит, что спрятанный объект никуда не исчезает. Постепенно узнает о существовании гравитации, инерции и жесткости — это базовые свойства, которые постигаются в основном путем наблюдения. У младенца нет средств воздействия на мир, но он много наблюдает и получает при этом огромное количество информации. Это делают и детеныши животных, но больше следуют инстинктам. Пока мы не выясним, как провести такое обучение, к созданию сильного ИИ не приблизимся. Есть и технические подзадачи, в которые я не буду углубляться, например предсказание в условиях неопределенности, но они уже вторичны.

М. Ф.: Но вы считаете, что сильный ИИ достижим?

Я. Л.: Конечно.

М. Ф.: И он обязательно появится?

Я. Л.: Я в этом не сомневаюсь.

М. Ф.: У него будет сознание или же это будет зомби, не имеющий сознательного опыта?

Я. Л.: Мы понятия не имеем, что такое сознание. Более того, я считаю, что в итоге вопрос наличия сознания окажется непринципиальным. Еще в XVII в., когда люди поняли, что на сетчатке глаза формируется перевернутое изображение, они были озадачены тем, что мы видим все неперевернутым. Когда поняли, как именно обрабатывается картинка, оказалось, что на самом деле не имеет значения, в каком порядке идут пикселы. Здесь то же самое. Я считаю сознание субъективным опытом, который появляется как побочный продукт интеллекта.

Есть несколько гипотез о том, как возникает иллюзия наличия сознания. По крайней мере, я считаю это иллюзией. Например, в префронтальной коре мозга есть механизм, позволяющий людям моделировать мир. Заметив какую-то ситуацию, мы подстраиваем под нее модель мира. Сознательное состояние — это своего рода форма внимания. Если бы наш мозг был больше и имел набор из нескольких механизмов для моделирования мира, мы обладали бы другим сознанием.

М. Ф.: Давайте поговорим о том, какие опасности несет ИИ. Считаете ли вы, что мы на пороге экономического спада с массовым исчезновением рабочих мест?

Я. Л.: Эти вопросы меня тоже интересуют, хотя я и не экономист. Экономисты называют ИИ технологией общего назначения (general-purpose technology, GPT). С их точки зрения эта технология затронет всю экономику, как электричество или паровой двигатель.

Меня беспокоит проблема безработицы. Технологии развиваются быстрее, чем у населения появляются навыки, необходимые

в новой экономической модели. Но экономисты утверждают, что скорость распространения технологий фактически ограничена долей людей, которые не умеют ими пользоваться. Другими словами, имеет место саморегуляция: чем больше людей остается не у дел, тем медленнее распространяются технологии. Вспомните, как обстояли дела с компьютерными технологиями.

М. Ф.: Но можем ли мы опираться на исторические случаи, рассматривая машины, у которых будут когнитивные способности? Мне кажется, на этот раз человечеству грозит более серьезный кризис.

Я. Л.: Я не думаю, что появление ИИ приведет к массовой безработице. Конечно, сейчас экономический ландшафт сильно изменился. Сто лет назад большая часть населения работала на полях, а теперь этим занимается 2 %. Как сказал один экономист, «перед нами всегда будут проблемы, которые нужно решать, а значит, работа не закончится». Грядущие ИИ-системы усилят человеческий интеллект так же, как механические машины увеличили физическую силу.

М. Ф.: Думаю, для водителей или работников фастфуда переход получится болезненным.

Я. Л.: Надо учитывать, что ценность товаров и услуг тоже поменяется. Все, произведенное машиной, станет дешевле, а сделанное людьми — дороже.

Например, проигрыватель Blu-Ray можно купить за 46 долларов. Если подумать о том, какая сложная технология лежит в его основе, цена кажется безумно низкой. Это «синие» лазеры, которых не было 20 лет назад. Невероятно точный сервомеханизм, обеспечивающий позиционирование лазера с точностью до микрона. Сжатие видео H.264 и сверхбыстрые процессоры. Но он стоит 46 долларов, потому что в основном производится машинами. Теперь зайдите в интернет и найдите керамическую салатницу

ручной работы. Технологии производства керамики 10 тысяч лет, но по первым же ссылкам вы найдете товар, который стоит примерно 500 долларов.

Поездки на такси будут дешевыми, потому что машиной управляет ИИ-система, но повысятся цены на еду в ресторанах, где готовит повар, а обслуживают официанты.

М. Ф.: Но ведь далеко не все люди обладают востребованными навыками или талантами. Что вы думаете об идее универсального базового дохода?

Я. Л.: У меня нет определенного мнения по этому поводу, так как я не экономист, но все экономисты, с которыми я разговаривал, выступали против универсального базового дохода. Все они считали, что правительства должны принять меры для компенсации возрастающего в результате технического прогресса неравенства. Все должно упираться в фискальную политику, то есть налогообложение, перераспределение богатства и доходов.

Неравенство в доходах особенно заметно в США, в меньшей степени — в Западной Европе. Коэффициент Джини — статистический показатель степени расслоения общества — во Франции или Скандинавии составляет около 25 или 30. В США он равен 45, и это уровень стран третьего мира. Экономист из MIT Эрик Бринолфссон вместе со своим коллегой Эндрю Макафи написал пару книг о влиянии технологий на экономику. Они говорят, что средний доход американских домохозяйств не изменился с 1980-х гг., то есть со времен рейганомики и снижения налогов на высокие доходы. А вот производительность росла более или менее непрерывно. В Западной Европе ничего подобного не было. Так что все зависит от налоговой политики. Существуют простые меры, которыми правительства могут компенсировать дестабилизацию, просто в США этого не делается.

М. Ф.: Какие еще опасности может принести ИИ?

Я. Л.: Начну с того, что нет смысла бояться, что мы создадим интеллект человеческого уровня, а он вырвется из-под нашего контроля, и роботы захотят захватить мир. Желание захватить мир связано не с интеллектом, а с тестостероном. В американской политике много примеров, явно показывающих, что стремление к власти не связано с интеллектом.

М. Ф.: Но, например, с точки зрения Ника Бострома проблема в том, что ИИ может начать добиваться своей цели способом, который навредит людям.

Я. Л.: То есть мы достаточно умны, чтобы создавать машины с интеллектом, а потом зачем-то приказываем им создать бесконечное количество скрепок? С моей точки зрения это нереально.

М. Ф.: Я думаю, что Ник специально привел преувеличенно карикатурный пример. Такие сценарии кажутся надуманными, но, когда речь заходит о суперинтеллекте, нельзя исключать, что машина может начать действовать непостижимыми для людей способами.

Я. Л.: Такие сценарии предполагают, что каким-то образом мы заранее проектируем целевую функцию, то есть внутренние мотивы машин, и в случае ошибки они начнут делать сумасшедшие вещи. Люди устроены по-другому. Внутренние целевые функции в нас не встроены. Разумеется, есть инстинкты, заставляющие есть, дышать и размножаться, но многое в нашем поведении и системе ценностей появляется как следствие обучения. То же самое можно сделать с машинами. Сформировать их систему ценностей, научить вести себя в обществе и приносить пользу человечеству. Не проектированием, а обучением, что гораздо проще. Мы же учим этому детей. Что мешает так же поступить с роботами или системами ИИ?

Мне эти вопросы кажутся преждевременными. Грубо говоря, у нас еще не изобретен двигатель внутреннего сгорания, а мы уже

беспокоимся, что не сможем изобрести тормоз и ремень безопасности. При этом изобрести двигатель куда сложнее.

М. Ф.: Что вы думаете о сценарии с рекурсивными улучшениями, в результате которых ИИ многократно превзойдет людей?

Я. Л.: Я в него не верю. Очевидно, что непрерывное улучшение неизбежно. Но чем выше будет интеллект машин, тем больше они начнут помогать в проектировании следующего поколения.

Существует дифференциальное уравнение, описывающее развитие технологий, экономику, потребление ресурсов, коммуникации, сложность технологий и т. п. Оно содержит целый ряд замедляющих коэффициентов, которые почему-то игнорируются сторонниками сценария быстрого взлета. Любой физический процесс в какой-то момент должен перейти в стадию насыщения, например по причине истощения ресурсов. Не будет такого, что кто-то раскроет секрет сильного ИИ и мы начнем переходить от машин с интеллектом крысы к машинам с интеллектом орангутана, через неделю они станут умнее нас, а месяц спустя — намного умнее.

Кроме того, превосходство в области интеллекта не означает превосходства во всем. Вирусы при всей своей глупости убивают людей. Если мы сможем создать систему ИИ, то, вероятно, сможем создать и более специализированный ИИ, предназначенный для уничтожения первого. И он будет крайне эффективно справляться с задачей, так как специализированные машины эффективнее универсальных. Я уверен, что в каждую проблему уже встроено решение.

М. Ф.: Тогда о чем имеет смысл беспокоиться в ближайшие десятилетия?

Я. Л.: Прежде всего о проблемах в экономике. Все они решаемы, но нужно преодолеть политические препятствия, особенно в таких

странах, как США, где не принимается идея о перераспределении доходов. Еще есть проблема с распространением технологии. Желательно, чтобы она приносила пользу всему миру, а не только развитым странам. Я думаю, что ускорение технического прогресса и появление ИИ заставят правительства больше вкладывать в образование, особенно в непрерывное образование, потому что людям придется искать новые рабочие места. Иначе нас ждет дестабилизация. Предвзятость, которая попадает в машину на уровне данных, в случае обучения с учителем может усугубиться.

М. Ф.: Я надеюсь, что предвзятость в алгоритме исправить проще, чем в человеке.

Я. Л.: Именно поэтому я смотрю на ситуацию с оптимизмом. Бороться с предвзятостью машин действительно проще. У людей необъективность и пристрастность иногда принимают твердую форму.

М. Ф.: Беспокоит ли вас применение ИИ военными, например создание автономного оружия?

Я. Л.: И да, и нет. Разумеется, технологию ИИ можно использовать для создания оружия, но, к примеру, я совершенно не согласен со Стюартом Расселом, который считает, что это непременно будет оружие массового уничтожения. Я думаю, что все будет с точностью до наоборот. ИИ начнут применять для так называемых хирургических действий. Зачем разрушать бомбой целое здание, если можно послать беспилотник, который просто усыпит человека, которого нужно захватить?

М. Ф.: Беспокоит ли вас то, что ИИ может первым создать Китай? В этой стране больше данных при меньших ограничениях на конфиденциальность. Дает ли им это преимущество?

Я. Л.: Не думаю. Современный прогресс в науке обусловлен не широкой доступностью данных. Сколько бы человек ни прожи-

вало в Китае, доля тех, кто связан с технологиями и занимается исследованиями, относительно мала.

Разумеется, их количество будет расти. Китай делает успехи в этом направлении. Но мне кажется, что принятые там стиль управления и тип образования могут через некоторое время задушить творческие порывы. Но в Китае есть и хорошие специалисты, которые могут внести вклад в развитие ИИ.

В 1980-х гг. Запад точно так же испытывал страх перед быстро распространяющимися японскими технологиями. Но в какой-то момент все стабилизировалось. Потом мы боялись корейцев, теперь вот китайцев. В течение следующих десятилетий в китайском обществе могут произойти большие перемены, что может полностью изменить ситуацию.

М. Ф.: Нуждается ли ИИ в регулировании со стороны государства?

Я. Л.: Я не думаю, что государство должно вмешиваться в исследования и процесс разработки ИИ, а вот конкретные варианты применения этой технологии, безусловно, потребуют регулирования. Но это требование связано не с ИИ, а с областью применения приложений. Представьте, что ИИ применяется для разработки лекарств. Процесс тестирования, выпуска и применения препаратов, содержащих наркотические вещества, уже регулируется. Когда на дорогах появятся беспилотные автомобили, они будут подчиняться ПДД. Разумеется, могут появиться и сферы, в которых все настолько изменится благодаря ИИ, что придется дорабатывать существующие правила. Но пока никакого регулирования в связи с ИИ не требуется.

М. Ф.: То есть вы не согласны с риторикой Илона Маска?

Я. Л.: Совершенно не согласен. Я несколько раз общался с ним, и не знаю, откуда взялись его взгляды. Маск умный парень, мне очень нравятся некоторые его проекты, но я не понимаю его

мотивации. Он хочет спасти человечество, возможно, поэтому ему нужна еще одна экзистенциальная угроза. Он искренне обеспокоен, но пока никто из нас не смог его убедить, что сценарии в стиле Ника Бострома человечеству не грозят.

М. Ф.: Получается, вы оптимист? И верите в то, что ИИ принесет больше пользы, чем проблем?

Я. Л.: Именно так.

М. Ф.: Как, по вашему мнению, можно этому способствовать?

Я. Л.: Я надеюсь, мы придумаем способ обучать машины, как маленьких детей и животных. И собираюсь этим заниматься следующие несколько лет. Еще я надеюсь, что серьезный прорыв произойдет до того, как люди, финансирующие эти исследования, устанут, как это случалось раньше.

М. Ф.: Вы предупреждали, что слишком сильная шумиха может привести к очередной «зиме ИИ». Вы действительно считаете, что это возможно? Глубокое обучение уже встроено в бизнес-модели таких корпораций, как Google, Facebook, Amazon, Tencent. Трудно представить, что инвестиции в технологию прекратятся.

Я. Л.: Я не думаю, что нам грозит еще одна «зима ИИ» в том виде, как это было раньше. Именно потому, что появились приложения, которые приносят реальную прибыль.

Но огромное количество инвестиций делается в надежде, например, на появление беспилотных автомобилей в ближайшие пять лет или на революцию в медицинской визуализации. Думаю, очень скоро это заметно повлияет на медицину и здравоохранение, транспорт и доступ к информации.

Другой случай — виртуальные помощники. Сейчас они пишутся вручную, поэтому толку от них немного. Вы видели фильм «Она»? Среди всех научно-фантастических фильмов об ИИ это, наверное,

один из наименее смешных. Там хорошо показано, что может произойти, если у каждого будет личный помощник, обладающий интеллектом человека.

С моей точки зрения, благодаря развитию аппаратного обеспечения доступ к связанным с ИИ технологиям получат многие. Поэтому усилия следует направить на разработку дешевого оборудования, которое потребляет мало энергии и может поместиться в смартфоне или в пылесосе, способном запустить сверточную сеть от мощности в 100 мВт. Если нужный чип можно будет купить за 3 доллара, в мире многое изменится.

Прогресс в данной области может оказать влияние и на экологию, обеспечив возможность мониторинга дикой природы. Благодаря развитию аппаратных технологий, предназначенных для глубокого обучения, в ближайшие два или три года ИИ станет доступен всем вокруг.

> *Посмотрите на специалистов в сфере ИИ — команды в компаниях, профессоров в академических кругах, аспирантов или докладчиков на ведущих конференциях: среди них очень мало женщин и представителей меньшинств.*

Фей-Фей Ли

ПРОФЕССОР COMPUTER SCIENCE И ДИРЕКТОР СТЭНФОРДСКОЙ ЛАБОРАТОРИИ ИИ (SEIL), ГЛАВА ОТДЕЛА ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ КОМПАНИИ GOOGLE, СОУЧРЕДИТЕЛЬ AI4ALL[1]

Фей-Фей Ли специализируется в областях компьютерного зрения и когнитивной нейробиологии. На основе принципов функционирования человеческого мозга разрабатывает алгоритмы, формирующие у компьютеров и роботов способность описывать увиденное. Ведет работу по расширению многообразия в сфере ИИ. Степень бакалавра физики получила в Принстоне, а степень доктора наук по электротехнике — в Калтехе.

[1] Дополнительные сведения об AI4ALL на сайте http://ai-4-all.org/

Мартин Форд: Как вы впервые заинтересовались ИИ и как складывался ваш карьерный путь?

Фей-Фей Ли: Меня всегда интересовали точные науки, в особенности физика. Поэтому я поступила в Принстонский университет, где в процессе обучения меня очаровало устройство Вселенной. Дальше мной двигало обычное любопытство. Я узнала, что Альберт Эйнштейн и Дэвид Шенберг — основатели современной физики — к концу жизни заинтересовались биологией и фундаментальными вопросами бытия. Мне тоже стало интересно не столько постигать физическую природу вещей, сколько понять, что представляет собой разум, благодаря которому люди заняли главенствующее положение на планете.

М. Ф.: Вы тогда еще жили в Китае?

Ф. Л.: Нет, мой интерес к ИИ и нейробиологии пробудился уже в США. К счастью, диссертацию на соискание докторской степени я писала по теме, находящейся на стыке этих двух дисциплин.

М. Ф.: Вы считаете, что изучение обеих областей дает преимущество и не стоит сосредоточиваться исключительно на подходах, ориентированных на computer science?

Ф. Л.: Это позволяет иметь свой взгляд на происходящее. Когнитивная нейробиология помогает мне рассматривать процессы с точки зрения алгоритмов и детальных моделей, связывать машинное обучение и процессы, происходящие в мозге человека. Ведь прогресс в сфере ИИ возникает именно благодаря попыткам повторить тот путь решения задач, который естественный интеллект прошел в ходе эволюции. Это уникальный подход к работе с ИИ.

М. Ф.: Вы высказали гипотезу, что в эволюционном плане развитие глаза, вероятно, привело к развитию мозга, который предоставлял вычислительную мощность для интерпретации изображений. Поэтому, возможно, понимание видения — это путь к пониманию интеллекта. Я прав?

Ф. Л.: Да. Важной частью человеческого интеллекта является язык: наряду с речью, тактильным осознанием, принятием решений и рассуждением. Но во все эти вещи встроено зрительное мышление. Природа спроектировала наш мозг так, что интеллект тесно связан с двигательной системой, принятием решений, эмоциями, намерением и языком. Мозг не только распознает изолированные объекты. Отвечающие за распознавание функции — неотъемлемая часть человеческого интеллекта.

М. Ф.: Можете кратко описать, что вы сделали для разработки машинного зрения?

Ф. Л.: В 2000-х гг. стояла цель научить компьютеры распознавать объекты. Ведь это умение позволяет людям ориентироваться в мире, понимать, что вокруг происходит, рассказывать о мире друг другу и т. д. В то время основным инструментом в области компьютерного зрения было машинное обучение.

Я окончила аспирантуру, занялась преподаванием и увидела, что модели на базе машинного обучения не дают нужных результатов. В то время международное сообщество занималось задачей по распознаванию 20 классов объектов — этого было недостаточно.

Меня в то время очень интересовал процесс развития когнитивных навыков. Мозг любого ребенка за первые несколько лет жизни обрабатывает огромное количество данных. Дети активно экспериментируют с окружающим миром, наблюдают за ним и таким образом постигают его. Как раз тогда началось бурное развитие интернета и появился доступ к большим объемам данных.

Мне в голову пришла идея все фотографии из сети распределить в соответствии со значимыми для людей концепциями и промаркировать. Результатом стал проект ImageNet с 15 млн аннотированных изображений. Мы с коллегами открыли доступ к базе данных ImageNet всему миру и начали проводить международные конкурсы для исследователей.

Поворотным стал 2012 год. Победитель конкурса ImageNet создал алгоритм, скомбинировав нашу базу данных, вычислительные мощности графического процессора и сверточные нейронные сети. Джеффри Хинтон написал статью, которая для меня стала первым шагом на пути к распознаванию объектов.

М. Ф.: Вы продолжаете работать над этим проектом?

Ф. Л.: Следующие два года мы совершенствовали процесс распознавания. Если посмотреть на стадии развития речевых навыков, младенцы сначала лепечут, потом произносят отдельные слова, а затем начинают говорить предложениями. Моя двухлетняя дочь уже говорит предложениями, и становится заметным прогресс в ее миропонимании. Мы хотим научить компьютеры реагировать на демонстрируемые изображения предложениями, а не просто находить присутствующие там объекты.

Мы работали над этой проблемой несколько лет, применяя модели глубокого обучения. В 2015 г. я сделала на конференции TED 2015 доклад «Как мы учим компьютеры понимать изображения».

М. Ф.: Но ведь это сильно отличается от того, что происходит с детьми. Ребенок наблюдает. Даже когда взрослый показывает ему маркированное изображение, достаточно сделать это несколько, но не сто тысяч раз. Обучение человека на неструктурированных, непрерывно поступающих данных и обучение с учителем ИИ-системы не получается поставить на одну плоскость.

Ф. Л.: Вы правильно поняли суть проблемы. Тот успех нейронных сетей и глубокого обучения, которого мы уже добились, это лишь небольшая часть возможностей интеллекта.

В этом году на конференции Google I/O я снова использовала в качестве примера свою дочь. Пару месяцев назад с помощью радионяни я наблюдала, как она ищет способы выбраться из кроватки. Я видела, как она открыла свой спальный мешок,

хотя он был специально сшит таким образом, чтобы ребенок не мог из него выбраться. Современные ИИ-системы не обладают такого рода скоординированным интеллектом, отвечающим за визуально-моторные навыки, планирование, мышление, эмоции, намерения и настойчивость. Так что нам предстоит еще много работы.

М. Ф.: Возможен ли прорыв, который позволит компьютерам учиться тем же способом, что и дети?

Ф. Л.: Над этим работает множество людей. В SEIL мы пытаемся заставить ИИ-системы обучаться путем подражания, что куда естественнее обучения с учителем. Поэтому начинаем применять алгоритмы обучения с подкреплением без прямого вознаграждения (inverse reinforcement learning, IRL) и алгоритмы нейропрограммирования. Этими исследованиями занимаются компания DeepMind, Google, мы и MIT.

Я не могу назвать дату возможного прорыва, потому что зачастую это дело счастливой случайности, когда внезапно совпадает множество различных факторов. Но надеюсь, что благодаря глобальным инвестициям в эту сферу все произойдет еще при нашей жизни.

М. Ф.: Выступая с презентациями, я всегда подчеркиваю, что со временем ИИ и машинное обучение станут вещами общего пользования, почти как электричество. И первым шагом к этому, на мой взгляд, является добавление ИИ в облачные сервисы. Как глава отдела облачных вычислений Google вы со мной согласны?

Ф. Л.: Именно поэтому я и оказалась в Google. У университетских профессоров, к счастью, есть возможность раз в семь или восемь лет брать творческий отпуск, и два года назад я присоединилась к индустрии, которая демократизирует технологии ИИ. Облако — самая лучшая и большая платформа для распространения технологий. Ведь сервисы Google Cloud в любой момент расширяют возможности миллиардов людей.

Например, мы занимаемся автоматическим обучением машин (AutoML). Это уникальный продукт, позволяющий неспециалистам пользоваться возможностями ИИ. Многим компаниям нужны индивидуальные модели: журналу National Geographic — модель для распознавания диких животных, а фирмам сельскохозяйственной отрасли — модель для распознавания овощей и фруктов. При этом у сотрудников этих фирм нет опыта в работе с ИИ, и они не могут самостоятельно выбрать наиболее подходящий алгоритм и оптимальные параметры.

М. Ф.: Похоже, что доступность машинного обучения, которое обеспечивает AutoML, может привести к появлению множества приложений ИИ, созданных разными людьми с разными целями.

Ф. Л.: Именно так! В своих презентациях я использую в качестве аналогии кембрийский взрыв.

М. Ф.: Сегодня нейронным сетям и глубокому обучению уделяется огромное внимание. Как вы считаете, это именно та технология, которая приведет к развитию ИИ? Или пора искать новые пути?

Ф. Л.: Если посмотреть на прогресс науки в целом, вы увидите, что в прошлом то и дело приходилось отказываться от каких-то вещей и даже отступать назад. Невозможно быть уверенными, что не появится новая, более совершенная методика. Это особенно верно для такой молодой сферы, как ИИ. Ведь она существует всего 16 лет.

М. Ф.: Какие проекты сейчас можно причислить к передовым исследованиям в области ИИ?

Ф. Л.: Моя лаборатория сейчас работает над проектом Visual Genome. В ImageNet связаны изображения и метки, тогда как в реальности существуют взаимосвязи между объектами, а также между зрением и языком. Поэтому проект Visual Genome можно назвать следующим шагом после ImageNet. Мы ищем связь между визуальным миром и человеческим языком.

Еще одно направление, в котором ИИ принесет пользу, это здравоохранение. Человеческий фактор сильно влияет на медицину: низкое качество обслуживания, отсутствие контроля, ошибки, высокие затраты, предвзятое отношение к пожилым людям. По этой теме вообще крайне мало доступной информации. Около пяти лет назад мы поняли, что технология на базе ИИ, предназначенная для внедрения беспилотных автомобилей, подходит для оказания медицинской помощи. Систему с датчиками для сбора информации об обстановке в больницах и настроении пациентов, алгоритмами для анализа собранных данных и обратной связи мы внедряем в Детской больнице Люсиль Паккард в Стэнфорде, Медицинском центре Intermountain в Юте и домах престарелых в Сан-Франциско.

М. Ф.: Какие препятствия нужно преодолеть для создания сильного ИИ?

Ф. Л.: Для начала следует определиться с термином. Лично для меня это интеллект, понимающий контекст ситуации и все детали, многогранный и многомерный, обладающий способностью к обучению не только на больших объемах маркированных данных, но и обучению с подкреплением и даже обучению без учителя.

Если отталкиваться от этого определения, получится, что нужно искать алгоритмы, выходящие за пределы обучения с учителем. Необходимо сотрудничество с нейробиологами, когнитивными психологами и специалистами по бихевиоризму, так как многие связанные с ИИ технологии находятся на стыке различных наук. В марте 2018 г. в газете *New York Times* была опубликована моя статья на эту тему *How to Make A.I.* («Как создать ИИ»).

М. Ф.: Я читал статью. Вы выступали за комплексный подход к следующей фазе разработки ИИ.

Ф. Л.: Все дело в том, что ИИ уже вышел за пределы академической среды и начал влиять на жизнь людей. Поэтому при его разработке и внедрении следует учитывать человеческий фактор.

Разработка ИИ должна стать междисциплинарной и ориентированной на человека. Сейчас много говорится о роботах, берущих на себя рутинную работу, но у ИИ есть и другие возможности улучшить качество жизни людей. Мне кажется, нужно инвестировать в технологии, касающиеся сотрудничества и взаимодействия людей и машин: робототехнику, обработку естественного языка и т. п.

М. Ф.: Ник Бостром, Илон Маск и Стивен Хокинг много говорили об экзистенциальной угрозе, которую несет рекурсивное самосовершенствование. Есть мнение, что одним из шагов к этому может стать ваш проект AutoML, ведь вы используете технологии для проектирования других систем машинного обучения.

Ф. Л.: Это здорово, что такие лидеры, как Ник Бостром, предупреждают о вещах, которые могут повлиять на людей неожиданным образом. Но важно учитывать контекст. Ведь любой новый социальный порядок или технология, которые появлялись в прошлом, имели разрушительный характер. Поэтому все это полезно изучать с разных сторон. Ник говорит о потенциальном влиянии ИИ на человечество как философ. Но я думаю, что в эту дискуссию должны внести свое слово и представители других специальностей.

М. Ф.: Такие люди, как Илон Маск, нагнетают всеобщую озабоченность, например, утверждая, что ИИ опаснее, чем Северная Корея. Он преувеличивает или у нас действительно есть повод для беспокойства?

Ф. Л.: Обычно запоминаются именно преувеличенные высказывания. Лично я предпочитаю прислушиваться к точкам зрения, основанным на веских доказательствах и логических выводах. Мне кажется, куда важнее то, как мы поступаем с теми возможностями и проблемами, которые актуальны сейчас. Поэтому меня больше волнуют вопросы предвзятости и отсутствие многообразия в сфере ИИ.

М. Ф.: То есть вы считаете, что экзистенциальная угроза — это дело далекого будущего?

Ф. Л.: Именно так. Но хорошо, что есть люди, которые думают об этом уже сейчас.

М. Ф.: Вы упоминали, что ИИ способен улучшить качество жизни людей, но у бизнеса есть мотив для сокращения рабочей силы. Это происходило на протяжении всей истории. Появятся ли в скором времени инструменты, способные автоматизировать не только рутинную, но и интеллектуальную работу? С вашей точки зрения, грозит ли нам массовая безработица и снижение заработной платы?

Ф. Л.: Капитализм — это одна из форм общественного устройства, которая существует порядка ста лет. И она не единственная. Предсказать, как технологии преобразуют наше общество, невозможно.

Говоря об улучшении жизни благодаря ИИ, я имею в виду увеличение продуктивности работы. За пять лет сотрудничества с врачами я убедилась, что часть их работы потенциально может выполняться машинами, освобождая их время для общения с пациентами и исследовательской работы, необходимой в случае редких или тяжелых заболеваний.

Технология ИИ обладает огромным потенциалом для совершенствования рабочей силы. Вспомните историю. Около 40 лет назад из-за появления компьютеров ушла в прошлое профессия машинистки. Но появились новые рабочие места. Теперь у нас есть инженеры-программисты. И это более интересная и творческая работа. А когда появились банкоматы и автоматизировали часть транзакций, в банке увеличилось количество служащих, так как возросло число доступных пользователям финансовых услуг.

М. Ф.: Вы много занимаетесь вопросами многообразия и предвзятости. Мне кажется, что эти вещи никак не пересекаются. Предвзятость может быть заложена в сгенерированных людьми

данных, на которых обучаются алгоритмы, тогда как многообразие — больше кадровый вопрос.

Ф. Л.: Эти темы не настолько далеки друг от друга, как вы думаете. Они связаны с ценностями, которые люди передают машинам. Многие ученые в сфере ИИ признают это явление и модифицируют алгоритмы, чтобы они имели возможность распознать такое смещение данных как предвзятость и исправить его.

М. Ф.: Каким образом в Google работают со смещениями данных для машинного обучения?

Ф. Л.: В Google над этим работает целая команда, поскольку существует ориентация на качество продукта. Я надеюсь, что в эту область будут делаться инвестиции. Что же касается темы многообразия и предвзятости людей… Это большая проблема, особенно в областях, связанных с точными науками. Посмотрите на специалистов в сфере ИИ — команды в компаниях, профессоров в академических кругах, аспирантов или докладчиков на ведущих конференциях: среди них очень мало женщин и представителей меньшинств.

М. Ф.: Расскажите о своем проекте AI4ALL, направленном на привлечение в связанные с ИИ сферы женщин и представителей меньшинств.

Ф. Л.: Именно недостаток многообразия четыре года назад побудил меня начать проект AI4ALL. Мы ориентируем старшеклассников в выборе профессии, и особенно нас интересуют представители меньшинств, потому что технология коснется всех. Это летняя учебная программа знакомства с ИИ. Инициатива имела такой успех, что в 2017 г. появилась национальная некоммерческая организация AI4ALL, к участию в которой стали приглашаться другие университеты.

Сейчас их шесть. Университет Беркли, например, предлагает программы для студентов с низким доходом, Принстонский

университет специализируется на программах для расовых меньшинств, Университет Кристофера Ньюпорта разрабатывает программы для трудных подростков, а Бостонский университет предлагает программы для девушек. Со временем мы надеемся на расширение.

М. Ф.: Как вы считаете, должно ли правительство взять в свои руки разработку правил для сферы ИИ, или профессиональное сообщество может самостоятельно решать возникающие проблемы?

Ф. Л.: Не думаю, что специалисты по ИИ смогут обеспечить общее благо. В нашем мире все тесно связано, и все мы зависим друг от друга. Я, будучи профессором, езжу по общим дорогам, дышу тем же воздухом и отправляю детей в школы. В работе над ИИ должны учитываться все сферы жизни. Правительство тоже играет огромную роль, инвестируя в фундаментальную науку, исследования и образование в области ИИ. Если нам нужна прозрачная и честная технология, если мы хотим, чтобы больше людей ее понимали и могли влиять на нее положительным образом, тогда помощь правительства необходима.

М. Ф.: Как вы относитесь к гонке вооружений в сфере ИИ? Например, Китай за счет авторитарной системы и численности населения располагает бóльшим объемом данных при меньшей степени конфиденциальности. Рискуем ли мы утратить свое лидерство в разработках ИИ?

Ф. Л.: В настоящее время достигнуты большие успехи в физике, и эти успехи влияют на развитие различных технологий. Кому принадлежит современная физика? Я считаю, что стремление человека к знаниям и истине не имеет границ.

Разумеется, конкуренция между компаниями и регионами существует, но я надеюсь, что и она во благо. Здоровая конкуренция означает уважение к соперникам, пользователям, рынку и законам. Я готова сотрудничать с людьми любого происхождения.

> *Работа над играми — это тренировка. Игры не являются конечной целью; мы хотим построить общие алгоритмы, которые можно будет применять к реальным задачам.*

Демис Хассабис

СОУЧРЕДИТЕЛЬ И ГЕНЕРАЛЬНЫЙ ДИРЕКТОР КОМПАНИИ DEEPMIND

Шахматный вундеркинд Демис Хассабис начал профессионально программировать и разрабатывать видеоигры в 16 лет. После окончания Кембриджского университета 10 лет основывал успешные стартапы, ориентированные на создание видеоигр и симуляций, основал компанию по производству видеоигр Elixir Studios. После получения докторской степени по когнитивной нейробиологии в Университетском колледже Лондона работал в MIT и Гарварде. Его исследование механизмов, лежащих в основе воображения и планирования, вошло в десятку лучших научных достижений 2007 г. по версии журнала Science.

Член Королевского общества искусств и Королевской инженерной академии, награжден серебряной медалью. В 2017 г. попал в список 100 самых влиятельных людей мира. Награжден орденом Британской империи за заслуги в науке и технике, получил премию Малларда и звание почетного доктора в Имперском колледже Лондона. В 2016 г. созданная в DeepMind программа AlphaGo победила чемпиона мира по игре в го Ли Седоля.

Мартин Форд: В юности вы увлекались шахматами и видеоиграми. Это повлияло на вашу карьеру в области ИИ и на решение основать DeepMind?

Демис Хассабис: С детства я мечтал стать чемпионом мира. Постоянно стремился улучшить свою игру, поэтому много думал о том, откуда мозг берет идеи для ходов. Какие процессы там происходят после хорошего хода или промаха? Позднее заинтересовался нейробиологией.

Шахматы много значат для ИИ. Идея компьютерной версии шахмат увлекала таких пионеров ИИ, как Алан Тьюринг и Клод Шеннон. В 8 лет на деньги, полученные за победу на шахматном турнире, я купил свой первый компьютер. И написал программу, имитирующую игру реверси. Это более простая игра, чем шахматы, но я использовал для нее те же идеи, например алгоритм альфа-бета-отсечения, которые первопроходцы ИИ применяли в своих шахматных программах. Фактически это была моя первая попытка написать программу с ИИ.

Следующим этапом стало написание коммерческих видеоигр. Одна из ключевых тем многих моих игр, от Theme Park (1994) до Republic: The Revolution (2003), — это симуляция. Игрокам предоставляется изолированная среда с персонажами, созданными на базе ИИ, реагирующими на различные действия.

Я считаю, что игры тренируют ум. Например, шахматы было бы полезно ввести в школьную программу, потому что они учат решать проблемы, планировать и в принципе прививают навыки, которые могут пригодиться в других областях. Возможно, все эти вещи я понимал подсознательно, когда основал DeepMind и начал использовать игры как среду обучения для ИИ-систем.

Перед этим я прослушал курс computer science в Кембриджском университете. Тогда, в начале 2000-х гг., мне не хватало идей для начала работы над сильным ИИ. В результате я получил степень

доктора нейробиологии, многое узнал о памяти и воображении — вещах, которые захотел перенести в машину.

М. Ф.: То есть вас с самого начала интересовал сильный ИИ?

Д. Х.: Именно так. Еще подростком я знал, чем хочу заниматься. Все началось с моего первого компьютера. Для меня это был волшебный инструмент: большинство машин дополняют физические возможности человека, а эта машина расширяла умственные способности.

Меня до сих пор восхищает, что для решения научной задачи можно написать программу, запустить ее и уйти спать. А утром получить готовый ответ. Естественно возникают мысли о следующем шаге: сделать машины, которые сами предлагают решения проблем.

М. Ф.: Компаний, которые специализируются именно на сильном ИИ, немного. Потому что нет бизнес-модели, позволяющей быстро получать доход. Как с этим справилась DeepMind?

Д. Х.: Мы с самого начала были компанией, нацеленной на разработку сильного ИИ, что осложнило поиск инвесторов. Наш тезис состоял в том, что технология общего назначения найдет сотни самых удивительных применений. Поэтому необходимость сначала собрать группу талантливых исследователей выглядела оправданной. В мире не так много людей, которые могли бы внести свой вклад в такую работу. В 2009–2010 гг., когда мы только начинали, можно было насчитать менее 100 человек. Вопрос был в том, сможем ли мы продемонстрировать четкий и измеримый прогресс.

В 2009 г. никакой шумихи вокруг ИИ не было. Более того, за последние 30 лет на эту тему было дано столько обещаний, которые окончились ничем, что получить первоначальное финансирование было невероятно трудно. Но у нас были сильные аргументы.

Во-первых, мы опирались на новые достижения в нейробиологии и понимании мозга. Во-вторых, собирались делать обучающиеся, а не традиционные экспертные системы. В-третьих, для ускорения разработки должно было применяться эталонное тестирование и симуляции. Мы смогли объяснить, почему ИИ не улучшался в предыдущие годы. Кроме того, наши методы работы требовали больших вычислительных мощностей, которые как раз в это время стали доступными.

В конце концов, нам удалось убедить достаточно людей, но направление, которое мы собирались разрабатывать, на тот момент было немодным. Его не одобряли даже в научных кругах, занимающихся ИИ, и вообще переименовали в «машинное обучение». Удивительно, как быстро все это изменилось.

М. Ф.: В итоге вы обеспечили финансирование и независимость компании. Почему вы продали ее Google?

Д. Х.: Мы не планировали продавать компанию, отчасти потому, что без продукта она не представляла ценности ни для одного крупного корпоративного бизнеса. У нас была бизнес-модель, которую мы не успели реализовать, и система компьютерного самообучения DQN (deep Q-network). К 2013 г. завершилась наша работа в Atari. Ларри Пейдж услышал о нас от одного инвестора, и мы неожиданно получили электронное письмо от вице-президента научного подразделения компании Google Алана Юстаса.

До момента продажи прошло некоторое время, потому что мне нужно было многое взвесить. Но в итоге я убедился, что ресурсы Google — их вычислительные мощности и возможность создать большую команду — ускорят выполнение нашей миссии по разработке сильного ИИ. Дело было не в деньгах: инвесторы были готовы увеличить финансирование.

Ларри и другие сотрудники Google также увлечены ИИ и предоставили нам автономию в том, что связано с направлениями ис-

следований и с нашей культурой; они согласились создать совет по этике, касающийся наших технологий. Кроме того, мы смогли остаться в Лондоне, что для меня было очень важно.

М. Ф.: Почему вы предпочли Лондон Кремниевой долине? Это связано с вами лично или с компанией?

Д. Х.: И со мной, и с компанией. Я родился и вырос в Лондоне и люблю этот город. Соседство Кембриджа и Оксфорда я считал конкурентным преимуществом. Причем тогда в Европе не было ни одной ставящей по-настоящему масштабные цели исследовательской компании, что давало нам высокие перспективы найма. К 2018 г. в Европе появилось несколько компаний, но мы были первыми, кто провел глубокие исследования в области ИИ. И мне кажется, что в таком деле должны принимать участие представители разных культур.

М. Ф.: Вы открываете лаборатории в европейских городах?

Д. Х.: Мы создали небольшую лабораторию в Париже, две лаборатории в Канаде — в Альберте и Монреале. После объединения с Google у нас появился офис в городе Маунтин-Вью, штат Калифорния.

М. Ф.: Насколько близко вы сотрудничаете с остальными ИИ-командами в Google?

Д. Х.: Над различными аспектами машинного обучения и ИИ в Google работают тысячи людей, которые занимаются как прикладными вопросами, так и исследованиями. Разумеется, все руководители групп знакомы друг с другом, и когда возникает такая необходимость, организуется сотрудничество. В отличие от остальных групп, DeepMind занимается исключительно сильным ИИ. У нас разработан долгосрочный план, базирующийся на данных о сути интеллекта и средствах его достижения, которые предоставляют нейробиологи.

М. Ф.: О вашей программе AlphaGo снят документальный фильм[1]. Думаю, она дает решения всем играм для двух игроков с открытой информацией. Планируете ли вы перейти к играм со скрытой информацией?

Д. Х.: Скоро выходит новая, улучшенная версия программы AlphaZero. Действительно, можно сказать, что мы разработали универсальное решение для игр типа шахмат, го, сеги и т. п. И пора делать следующий шаг. Сейчас мы работаем над стратегической игрой для ПК StarCraft со сложным игровым пространством. Там нет статичного набора фигур, как в шахматах, потому что игроки строят свои юниты. Кроме того, присутствует скрытая информация, так называемый «туман войны». Игрок не видит фрагментов экрана, пока не исследует эту область.

Работа над играми — это тренировка. Игры не являются конечной целью; мы хотим построить общие алгоритмы, которые можно будет применять к реальным задачам.

М. Ф.: До сих пор вы в основном сочетали глубокое обучение и обучение с подкреплением. Это правда, что вы считаете обучение с подкреплением способом достижения сильного ИИ?

Д. Х.: Да, это так. Это очень мощный метод, но его нужно объединять с другими. Обучение с подкреплением известно давно, но применялось оно только для решения модельных задач из-за трудностей масштабирования. Во время работы в Atari мы добавили к нему глубокое обучение, которое отвечало за обработку экрана и моделирование среды игры, и подошли к решению более крупных задач в программе AlphaGo и системе DQN. Все эти вещи лет десять назад считались невозможными.

Мы одна из немногих компаний, которые относятся к обучению с подкреплением серьезно, потому что основываемся на представ-

[1] https://www.alphagomovie.com/

лении о нем в нейробиологии. Речь идет о так называемом обучении на основе временных разностей, или TD-обучении (temporal difference learning). Оно реализуется благодаря системе выработки дофамина. Синтезирующие дофамин нейроны в случае ошибок снижают уровень его выработки, что заставляет в будущем избегать подобных ситуаций, то есть учиться на ошибках. В ответ же на положительные стимулы выработка дофамина увеличивается. Это принцип работы мозга — единственного известного нам примера интеллекта. Возможно, существуют и другие пути, но с точки зрения биологии кажется, что достаточно научиться масштабировать обучение с подкреплением.

М. Ф.: Но ведь когда ребенок учится говорить или познает мир, ни о каком обучении с подкреплением речи не идет. Это обучение без учителя — наблюдение или случайные взаимодействия с окружающей средой.

Д. Х.: Ребенок учится множеством способов: обучение с учителем реализуется при помощи родителей, учителей или сверстников, а экспериментируя с окружающим миром, дети учатся без учителя. Когда ребенок получает похвалу, это уже обучение с подкреплением. Мы работаем над всеми тремя вариантами. Обучение без учителя чрезвычайно важно. Вопрос в том, существует ли внутренняя, эволюционно заложенная мотивация, которая в конечном итоге обеспечивает вознаграждение при обучении без учителя? Есть доказательства того, что сам процесс получения информации воспринимается мозгом как вознаграждение. Имеет место также поиск новизны. Новые впечатления приводят к выработке дофамина.

М. Ф.: Я почувствовал, как глубоко вы интересуетесь нейробиологией и computer science. Сказывается ли это на подходах, которые используются в DeepMind?

Д. Х.: Да, я получил образование в обеих областях. В компании DeepMind больший упор делается на машинное обучение. При

этом самая большая группа, возглавляемая профессором Принстонского университета Мэттом Ботвинником, состоит из нейробиологов.

Проблема в том, что нейробиология — обширная область, и если специалист по машинному обучению обратится к ней по какому-либо вопросу, он просто запутается в огромном массиве информации. Многие говорят, что исследования ИИ базируются на нейробиологии, но не могут объяснить, как это происходит. Существуют две крайности. В проекте Blue Brain делаются попытки смоделировать мозг на уровне коры…

М. Ф.: Это проект, который возглавляет Генри Маркрам?

Д. Х.: Да. Там пытаются реконструировать колонки кортекса. Это может быть интересно с точки зрения нейробиологии, но, на мой взгляд, это не самый эффективный путь к созданию ИИ. Все происходит на слишком низком уровне. Мы же в DeepMind пытаемся понять мозг на уровне систем и алгоритмов, которые он реализует, и возможностей, функций и представлений, которые он использует. Нас не интересует точное устройство человеческого мозга. Нет никакой причины создавать компьютерную модель, точно копирующую, например, образование новых нейронов гиппокампа. Но очень интересно, каким способом реализуются функции, за которые он отвечает: эпизодическая память и ориентация в пространстве.

М. Ф.: Самолеты летают, как и птицы, но при этом им не приходится хлопать крыльями.

Д. Х.: Прекрасная аналогия. Да, можно сказать, что мы в DeepMind как бы пытаемся понять принципы аэродинамики, наблюдая за полетом птиц, чтобы потом абстрагироваться от деталей этого полета и создать самолет. До изобретения аэродинамического профиля были только безуспешные попытки использовать деформируемые крылья. Мы поняли, что мозг масштабирует

обучение с подкреплением, и ведем разработки в этом направлении. Важно научиться сужать пространство поиска. Этот момент часто упускают специалисты в области ИИ, игнорирующие нейробиологию.

М. Ф.: Недавно в DeepMind смоделировали нейроны решетки, отвечающие за пространственную ориентацию. Это случай, когда одна и та же базовая структура естественным образом возникает как в мозге, так и в искусственных нейронных сетях.

Д. Х.: Это одно из наших крупнейших достижений за последний год. Нам написали Эдвард и Мэй-Бритт Мозер, которые в свое время получили Нобелевскую премию за открытие нейронов решетки. Они предположили, что эти нейроны дают оптимальный способ представления пространства при вычислениях. Теперь нейробиологи проверяют, статичны ли эти нейроны или модифицируются в структуру на ходу, что лучше всего подходит для самообучающейся системы.

Кроме того, недавно на базе наших ИИ-алгоритмов мы создали новую теорию о том, как может работать префронтальная кора головного мозга. Я думаю, со временем работа ИИ-алгоритмов заставит нас по-другому посмотреть на устройство мозга, поскольку является хорошим аналитическим инструментом для экспериментов. Через сравнение ИИ-системы с человеческим мозгом можно изучать природу сознания, творчества и сновидений.

М. Ф.: Вы считаете, что есть общие принципы интеллекта, не зависящие от среды, в которой он возникает?

Д. Х.: Именно так. Определение этих общих принципов даст ключ к пониманию человеческого мозга.

М. Ф.: Каким образом ваши достижения смогут применить на практике в ближайшем будущем?

Д. Х.: Вы уже пользуетесь множеством приложений. Это и машинный перевод, и анализ изображений, и компьютерное зрение.

Компания DeepMind начала работу над такими вещами, как оптимизация энергии в центрах обработки данных Google. Система преобразования текста в речь WaveNet теперь есть в помощнике Google во всех телефонах на платформе Android. ИИ применяется в системах рекомендаций, магазине Google Play и может в фоновом режиме экономить заряд аккумулятора в телефонах Android. Все эти вещи используются каждый день. И я думаю, что это только начало.

Надеюсь, через некоторое время мы начнем сотрудничать со сферой здравоохранения. Например, уже сейчас в известной британской офтальмологической больнице Moorfields мы диагностируем макулодистрофию по результатам сканирования сетчатки. Результаты первого этапа нашего партнерства опубликованы в журнале *Nature Medicine*. Они показывают, что наша ИИ-система интерпретирует результаты сканирования с беспрецедентной точностью. К тому же она дает рекомендации по лечению более 50 заболеваний глаз на уровне ведущих мировых экспертов. Существуют и другие команды, выполняющие аналогичную работу для таких заболеваний, как рак кожи. Думаю, что в течение следующих пяти лет наша работа сможет принести еще много пользы.

Но больше всего меня радует, что ИИ вот-вот начнут применять для решения научных проблем. Мы работаем над изучением механизма сворачивания белка, значит, сможем проектировать материалы и создавать лекарства. ИИ уже используют для анализа данных с Большого адронного коллайдера, для поиска экзопланет. Существует множество массивов данных, структуру которых экспертам определить трудно, и можно озадачить этим вопросом ИИ. Надеюсь, благодаря этому в следующем десятилетии нас ждут научные достижения в фундаментальных областях.

М. Ф.: Какие препятствия нужно преодолеть на пути к сильному ИИ?

Д. Х.: Эти вещи мы определили с самого начала работы DeepMind. Во-первых, это получение абстрактных, концептуальных знаний с последующим переносом обучения. Люди достаточно легко переносят свои знания из одной области в другую. Человек, получив задачу нового для себя типа, не начинает изобретать велосипед, потому что может использовать уже имеющиеся знания и опыт из других областей. Компьютерные системы в подобных случаях требуют множество данных и работают крайне неэффективно. Это нужно исправить.

Во-вторых, требуется улучшить понимание естественного языка, а также скопировать, используя новые техники, то, что умели делать старые системы ИИ, например символьные манипуляции. Прорывом станет решение именно этих проблем.

М. Ф.: Когда сильный ИИ появится, будет ли он обладать сознанием?

Д. Х.: Я надеюсь, что ответ на этот вопрос появится в процессе работы. Но с моей точки зрения, сознание и интеллект разобщены. Одно без другого вполне может существовать.

М. Ф.: Сильный ИИ может стать разумным зомби?

Д. Х.: Возникает философский вопрос, как об этом узнать, если машина ведет себя так же, как и мы? Если воспользоваться бритвой Оккама, получится, что если некто демонстрирует такое же поведение, как и я, и сделан из того же материала, то можно предположить, что он чувствует то же самое, что и я. Но машина сделана из другого материала. И в этом случае уже нельзя безоговорочно применить бритву Оккама. Может быть, в некотором смысле машины сознательны, но на уровне чувств мы этого не ощущаем, потому что они сильно отличаются от нас.

М. Ф.: То есть вы верите в возможность сознания у машин? Что это не биологический феномен.

Д. Х.: У меня нет подобных предубеждений. Пока мы не знаем ответа. Вполне может оказаться, что в биологических системах есть что-то особенное. Например, сторонник гипотезы квантового сознания сэр Роджер Пенроуз считает, что создание сильного ИИ невозможно. Но я надеюсь, что путь, по которому мы идем, даст понимание о том, что такое сознание и где оно находится.

М. Ф.: Какие риски, на ваш взгляд, связаны с сильным ИИ? Илон Маск говорил о «призывах демона» и экзистенциальной угрозе. Много об этом писал и Ник Бостром, который, насколько я знаю, входит в совет экспертов DeepMind. А что об этом думаете вы?

Д. Х.: Я много говорил с ними об этих вещах. И как это всегда бывает при личных беседах, проявились различные нюансы высказываний, которые изначально казались не тем, чем стали.

Лично я работаю над ИИ, потому что считаю, что это принесет пользу человечеству. Позволит раскрыть наш потенциал в науке. Любая технология сама по себе нейтральна. А будущее зависит от того, как мы, люди, решим ее использовать и распределить выгоды.

Сложностей много, но они больше напоминают геополитические проблемы, которые нужно решать обществу. Ник Бостром беспокоится о технических вопросах, таких как проблема контроля и проблема выравнивания. Мне кажется, что говорить об этом пока рано и нужны дополнительные исследования, ведь обсуждаемые системы появились недавно.

Пять лет назад все эти вопросы вообще были чисто философскими. Теперь у нас есть AlphaGo и зарождаются другие интересные технологии. На текущей стадии нужно делать обратное проектирование этих вещей и экспериментировать с ними, создавая

инструменты визуализации и анализа. Это позволит лучше понять, как функционируют эти «черные ящики» и как мы интерпретируем их поведение.

М. Ф.: Уверены ли вы, что мы справимся с опасностями, которые несет сильный ИИ?

Д. Х.: Когда эти системы перестанут быть «черным ящиком», мы разберемся, как ими управлять. Я уверен, что через инженерные эксперименты можно решить многие проблемы, о которых беспокоится Ник Бостром. Теория и практика должны идти рука об руку. Это не значит, что я не вижу поводов для беспокойства. Но я предпочитаю заниматься актуальными проблемами. Такими, как тестирование систем, которые мы внедряем в продукты. Конечно, некоторые из долгосрочных проблем настолько сложны, что возникает желание думать о них уже сейчас, задолго до того, как появится реальная необходимость искать их решение.

М. Ф.: А что вы думаете об автономном оружии? Вы очень откровенно высказывались об использовании ИИ в военных целях.

Д. Х.: В DeepMind мы считаем, что приложения ИИ должны оставаться под контролем людей и применяться в общественно полезных целях. То есть мы за запрет разработки и развертывания полностью автономного оружия и выражали эту точку зрения разными способами, в том числе подписали открытое письмо и поддержали обращение FHI.

М. Ф.: Но ведь химическое оружие все равно применяется, несмотря на запрет. Из-за соперничества между странами ситуация может начать развиваться в другом направлении. Например, в Китае активно работают над ИИ. Стоит ли волноваться о том, что они получат преимущество в этой сфере?

Д. Х.: Я не считаю, что имеет место гонка вооружений. Мы знаем всех исследователей и много сотрудничаем. Статьи публикуются

открыто, и, например, мне известно, что холдинг Tencent создал клон AlphaGo. Но если в будущем возникнет необходимость в координации и, возможно, даже регулировании, это должно быть на международном уровне и принято всеми. Иначе регулирование просто не сработает. Но эта проблема связана не только с ИИ. Есть множество других проблем, которые нужно решать на глобальном уровне, например проблему изменения климата.

М. Ф.: Как ИИ повлияет на экономическую ситуацию? Ждет ли нас рост безработицы и неравенства?

Д. Х.: До сих пор ИИ не очень сильно влиял на экономическую ситуацию, по сравнению с влиянием технического прогресса в целом. Но со временем ИИ многое изменит. Некоторые верят в новую промышленную революцию, сравнивая ИИ с электричеством. Но пока невозможно однозначно сказать, что нас ждет. Мы можем оказаться в мире изобилия благодаря росту производительности. Здесь главное, чтобы плодами прогресса смогли воспользоваться все. Пусть этими вещами занимаются экономисты.

М. Ф.: Да, мне тоже кажется, что проблема в основном сводится к распределению благ и что значительная часть населения окажется за бортом. Это сложнейшая политическая задача — создать новую экономическую парадигму, которая будет работать для всех.

Д. Х.: Именно так. Хотя рост производительности предсказывают уже в течение ста лет. Мой отец в университете изучал экономику. И рассказывал, что в конце 1960-х гг. многие всерьез задавались вопросом, что люди будут делать в 1980-х гг., когда наступит изобилие и исчезнет необходимость работать. Но мы работаем до сих пор, причем не менее усердно.

М. Ф.: То есть вас можно назвать оптимистом? Ведь вы считаете ИИ одной из лучших вещей, которые когда-либо случались с че-

ловечеством, при условии, что этой технологией сумеют разумно распорядиться.

Д. Х.: Именно поэтому я работаю над этим всю жизнь. Мне кажется, что не появись ИИ, мир развивался бы не в лучшую сторону. Разумеется, перед нами множество проблем, требующих решения, таких как изменение климата, болезнь Альцгеймера или очистка воды. Некоторые вещи со временем будут становиться только хуже. Я не понимаю, как мы сможем организовать глобальную координацию, получить избыточные ресурсы или придумать варианты решения актуальных проблем. Но в целом я оптимистично смотрю на будущее, потому что нас ждут такие технологии, как ИИ.

> *Обучение с учителем подарило, наверное, каждой крупной отрасли новые возможности, но я думаю, что мы сможем придумать еще более продуктивный подход.*

Эндрю Ын

ГЕНЕРАЛЬНЫЙ ДИРЕКТОР КОМПАНИИ LANDING AI, СОЗДАТЕЛЬ ПРОЕКТА AI FUND, ПРОФЕССОР COMPUTER SCIENCE, ЧЛЕН СОВЕТА ДИРЕКТОРОВ КОМПАНИИ DRIVE.AI

Эндрю Ын известен и как ученый, и как предприниматель. Именно он стоял у истоков проекта Google Brain и компании онлайн-образования Coursera. Затем стал ведущим специалистом в компании Baidu, где создал исследовательскую группу по ИИ. С 2018 г. он специализируется на стартапах в сфере ИИ. Запустил онлайн-платформу deeplearning.ai. В настоящее время занимается проектом Woebot. Был директором SAIL. Степень бакалавра в области computer science получил в Университете Карнеги — Меллона, степень магистра — в MIT и степень доктора — в Калифорнийском университете в Беркли.

Мартин Форд: Расскажите о перспективах развития ИИ. Глубокое обучение и дальше останется основным подходом или его заменит что-то другое?

Эндрю Ын: Пока всеми успехами в сфере ИИ мы обязаны обучению с учителем, которое в основном учит сопоставлять входные и выходные данные. Например, на вход беспилотного автомобиля подается видео с информацией о том, что находится впереди, а на выходе мы получаем сведения о фактическом положении других автомобилей. При распознавании речи машина сопоставляет аудиоклип и его текстовую расшифровку. В случае машинного перевода — текст на одном языке сопоставляется с текстом на другом.

М. Ф.: Какие препятствия отделяют нас от сильного ИИ?

Э. Ы.: Пока на этот вопрос нет однозначного ответа. Скорее всего, понадобится освоить обучение без учителя. Сегодня, чтобы компьютер узнал, что такое кружка, ему показывают тысячи кружек. В реальности такого не делает ни один, даже самый терпеливый и любящий, родитель. Дети просто смотрят на мир и погружаются в него. Что такое кружка, они узнают опытным путем. Именно такой вариант обучения коренным образом повысит эффективность наших систем.

Мы добились большого прогресса в специализированных вариантах ИИ для онлайн-рекламы, распознавания речи и беспилотных автомобилей, которые широкая публика ошибочно принимает за успехи в разработке сильного ИИ. В реальности подход к нему не найден. Неосведомленные люди пользуются упрощенными экстраполяциями, что создает ненужный ажиотаж вокруг этой темы.

М. Ф.: Появится ли сильный ИИ при вашей жизни?

Э. Ы.: Не знаю. Я бы этого очень хотел, но, скорее всего, потребуется намного больше времени.

М. Ф.: Как возник ваш интерес к ИИ? И как он повлиял на вашу карьеру?

Э. Ы.: На стажировке в школе я задумался, можно ли как-то автоматизировать часть моей работы, и именно тогда я узнал о нейронных сетях. Я учился и в итоге защитил докторскую диссертацию *Shaping and Policy Search in Reinforcement Learning* («Формирование и поиск стратегии в обучении с подкреплением») в Калифорнийском университете в Беркли. Следующие двенадцать лет я преподавал на факультетах computer science и электротехники Стэнфордского университета. В 2011 г. принимал участие в запуске проекта Google Brain, который помог превратить Google в компанию, занимающуюся ИИ.

М. Ф.: Это была первая попытка использовать глубокое обучение в компании Google?

Э. Ы.: В определенной степени да. У Google было несколько небольших проектов, связанных с нейронными сетями, но именно наша команда инициировала активное использование глубокого обучения. В качестве руководителя проекта я первым делом организовал курсы по этой теме для сотни инженеров. Это обеспечило команду Google Brain партнерами. Сначала мы взаимодействовали с группой распознавания речи. Позже поставили эксперимент с обучением без учителя, в котором нейронной сети предоставили множество случайных кадров из роликов YouTube, и она научилась распознавать кошек. В настоящее время толку от обучения без учителя не очень много, но эксперимент хорошо продемонстрировал, чего можно достичь с помощью алгоритмов глубокого обучения, используя вычислительный кластер Google.

М. Ф.: Вы работали в Google до 2012 г. А потом?

Э. Ы.: В какой-то момент я почувствовал, что для глубокого обучения лучше использовать графические процессоры. И перешел

в Стэнфорд. Я хорошо помню разговор на эту тему с Джеффри Хинтоном на ежегодной Конференции по машинному обучению и нейровычислениям. Думаю, наша беседа могла повлиять на работу Джеффри с Алексом Крижевским.

Мне повезло, что я преподавал в Стэнфорде именно в то время. Там как нигде чувствовалось начало эры неспециализированных вычислений на графических процессорах. Мы раньше других увидели, что на таких процессорах лучше масштабировать алгоритмы глубокого обучения. Мой бывший студент Адам Коутс доказал, что чем больше данных дается алгоритмам глубокого обучения, тем лучше они работают. И я попросил Ларри Пейджа дать добро на использование компьютеров для создания очень большой нейронной сети.

М. Ф.: После этого с Дафной Коллер вы начали работу над проектом Coursera, а потом перешли в корпорацию Baidu?

Э. Ы.: Я помогал Дафне начать проект Coursera, так как хотел, чтобы онлайн-обучение было доступно как можно большему числу людей. На тот момент команда Google Brain уже настолько хорошо функционировала, что я был рад передать бразды правления Джеффу Дину. И пару лет работал над проектом Coursera, который мы создавали с нуля. Это продолжалось до 2014 г., после чего я перешел работать в отдел AI Group компании Baidu. Подобно тому как проект Google Brain преобразовал Google в компанию, работающую на базе ИИ, наша группа помогла осуществить аналогичную трансформацию с Baidu. Через три года я решил пойти дальше и стал генеральным директором стартапа Landing AI и партнером стартапа AI Fund.

М. Ф.: Теперь вы хотите масштабировать и трансформировать все вокруг?

Э. Ы.: Да, крупные поисковые системы я уже преобразовал, пришло время других отраслей. Проект Landing AI призван помочь

компаниям применить возможности, которые открывает ИИ. Следующий шаг в этом направлении делает проект AI Fund, который ищет способы создания новых предприятий с использованием ИИ-технологий. Перед нами постоянно открываются новые перспективы. Пару десятков лет назад Google, Amazon, Facebook и Baidu по большому счету были стартапами, но все эти компании шли в ногу со временем. Проект AI Fund призван поддержать новые стартапы, которые собираются работать на базе ИИ.

М. Ф.: Многие считают, что имеющие доступ к огромному количеству данных гиганты, вроде компаний Google и Baidu, не оставляют шанса новым компаниям. Вы считаете, что у стартапов есть перспективы?

Э. Ы.: Разумеется, активы в виде накопленных данных, которыми обладают крупные поисковые системы, вряд ли оставляют шансы для новичков. Но при этом непонятно, насколько данные о потоках кликов могут пригодиться для медицинской диагностики, производства или индивидуального образования. Я думаю, что существует большая потребность в специализированных данных, и вокруг этого можно строить бизнес.

М. Ф.: Чем ваш проект AI Fund отличается от множества других венчурных компаний?

Э. Ы.: Венчурные фонды пытаются определить, насколько успешным может стать предлагаемый вариант бизнеса, и на этой основе принимают решения. Мы же создаем успешные бизнесы. Наша специализация — построение с нуля. Мы работаем с совершенно новыми командами, наставляем и поддерживаем их. От тех, кто заинтересован в работе с нами, не требуется исчерпывающей информации, достаточно резюме. Идея стартапа реализуется после начала совместной работы.

М. Ф.: К вам приходят люди с готовыми идеями или вы всегда помогаете с разработкой концепции?

Э. Ы.: Если человек принесет свою идею, мы с удовольствием ее обсудим, но у нас есть длинный список многообещающих идей, для которых нужны исполнители. И мы всегда рады ими поделиться.

М. Ф.: То есть вы привлекаете талантливых людей, предлагая возможность и инфраструктуру для стартапа?

Э. Ы.: Именно так. Хотя, конечно, для создания успешной компании требуется не только талант в сфере ИИ. Технологии уделяется так много внимания, потому что она очень быстро развивается. Но для формирования сильной команды нужен целый набор навыков: знание технологии, умение строить бизнес-стратегии, разрабатывать и продвигать продукт и развиваться. Мы создаем вертикально интегрированные компании.

М. Ф.: Кажется, любой стартап в сфере ИИ, демонстрирующий реальный потенциал, приобретается одним из технологических гигантов. Есть ли у современных стартапов шансы дойти хотя бы до первичного публичного предложения?

Э. Ы.: Я очень надеюсь, что какие-то ИИ-стартапы смогут продолжить самостоятельное существование. На самом деле мы не ставим перед собой финансовых целей, хотелось бы просто сделать что-то полезное. И грустно думать, что руководители всех успешных стартапов рано или поздно продадут свое детище.

М. Ф.: В последнее время все чаще идут разговоры о том, что успехи глубокого обучения преувеличены. Говорят даже о новой «зиме ИИ». Насколько это соответствует реальности и есть ли риск уменьшения инвестиций в эту сферу?

Э. Ы.: Я сильно сомневаюсь в наступлении очередной «зимы ИИ», но снизить ожидания относительно сильного ИИ однозначно стоит. Предыдущие «зимы» сопровождались ажиотажем по поводу новых технологий, которые не были такими уж полезными и принесли меньшую выгоду, чем ожидалось. Рост количества проектов, специалистов и компаний, работающих с глубоким обучением, означает,

что сегодня эта сфера приносит доход. Инвестиции в глубокое обучение идут непрерывно. Поддержка со стороны крупных компаний базируется не на надеждах и мечтах, а на достигнутых результатах.

М. Ф.: То есть если не брать в расчет ожидания относительно сильного ИИ, вы думаете, что нас ждет прогресс в сфере глубокого обучения и новые специализированные приложения на базе этой технологии?

Э. Ы.: Современный ИИ сильно ограничен. Кроме того, сам термин описывает очень широкий набор понятий, и я думаю, что при обсуждениях ИИ в большинстве случаев подразумеваются такие инструменты, как метод обратного распространения, обучение с учителем и нейронные сети.

Он ограничен так же, как интернет или электричество. От того, что доступ к электричеству превратился в коммунальную услугу, проблемы человечества не исчезли. Не стоит ждать этого и от метода обратного распространения, несмотря на всю его эффективность. Но ясно, что от нейронных сетей, обученных с помощью этого метода, мы получили далеко не всё.

Иногда мои выступления на тему ИИ начинаются с фразы: «ИИ не волшебная палочка, он не может делать все». Странно, что в современном мире до сих пор кто-то верит во всемогущество технологий. Мы достигли огромных успехов в узком ИИ, и сильно продвинулись в общем ИИ, но обе эти вещи обозначаются одним термином. В итоге экономические блага, которые удалось получить благодаря узкому ИИ, формируют ошибочное мнение о прогрессе в общем или даже сильном ИИ. А здесь пока хвастаться нечем.

М. Ф.: Смогут ли нейронные сети обеспечить постоянный прогресс в сфере ИИ или же потребуется гибридный подход, включающий идеи из других областей, например символической логики?

Э. Ы.: В компании Landing AI при создании решений для промышленных партнеров постоянно используются гибридные

инструменты. При работе с небольшими наборами данных часто приходится комбинировать методики глубокого обучения, например, с традиционными инструментами компьютерного зрения. Специалист в сфере ИИ должен понимать, в каких случаях лучше воспользоваться набором методов и как их оптимальным образом скомбинировать. Это что касается быстрой разработки полезных приложений.

Но если смотреть в долгосрочной перспективе и предположить, что мы все-таки сможем приблизиться к интеллекту уровня человека, скорее всего, произойдет сдвиг в сторону нейронных сетей. Мне кажется, что будут придуманы алгоритмы, намного превосходящие метод обратного распространения.

М. Ф.: То есть вы считаете нейронные сети лучшей технологией для прогресса в сфере ИИ?

Э. Ы.: Я думаю, что в обозримом будущем центральное место займут именно они. Достойной альтернативы нейронным сетям я пока не вижу, но это не значит, что она никогда не появится.

М. Ф.: Джуда Перл убежден, что для прогресса в сфере ИИ нужна модель, учитывающая причинно-следственные связи. Вы с этим согласны?

Э. Ы.: Есть множество вещей, которых глубокое обучение не делает. Учет причинности — одна из них. Кроме того, нужно учиться работать с небольшими наборами данных, совершенствовать многозадачное обучение, отказаться от меток. Как видите, у метода обратного распространения много недостатков, которые не мешают продвигаться вперед, но над их устранением нужно работать.

М. Ф.: Я видел исследование, показывающее, что сети глубокого обучения легко обмануть с помощью сфабрикованных данных.

Э. Ы.: Да, и это большая проблема. В Baidu мы постоянно боролись с атаками на ИИ-системы и с попытками мошенничества.

М. Ф.: Что вы думаете о проблеме конфиденциальности? Например, в Китае сейчас повсеместно начинает внедряться технология распознавания лиц.

Э. Ы.: Я не эксперт в этой области, поэтому просто озвучу наиболее близкую мне точку зрения. Развитие технологий дает потенциал для концентрации влияния. Так было с интернетом. Власть оказывается у корпораций или правительства, и небольшая группа людей получает возможность влиять на остальных.

Современные технологии позволяют, к примеру, влиять на результаты голосования, что заставляет беспокоиться о будущем демократии. Конечно, на недавних выборах в США применялись в основном интернет-технологии, а до этого огромное влияние оказывало телевидение. Поэтому нужна постоянная готовность противодействовать злоупотреблениям.

М. Ф.: Расскажите об одном из самых популярных приложений ИИ: беспилотных автомобилях. Когда, по вашему мнению, эта услуга станет общедоступной?

Э. Ы.: Я думаю, что в геозонах беспилотные автомобили появятся сравнительно быстро, возможно, уже к концу этого года, но на дорогах общего пользования вы их увидите нескоро.

М. Ф.: Под геозоной вы подразумеваете заранее нанесенные на карту маршруты?

Э. Ы.: Да. Некоторое время назад мы с коллегой написали статью для журнала *Wired*[1], в которой предложили вариант внедрения беспилотных автомобилей. Для их массового использования требуются как инфраструктурные, так и социально-правовые изменения.

[1] https://www.wired.com/2016/03/self-driving-cars-wont-work-change-roads-attitudes/

Мне посчастливилось более 20 лет наблюдать, как развивалась эта индустрия. Еще студентом я видел, как Дин Померло обучал автономный автомобиль, в котором был установлен следивший за дорогой компьютер. Но эта великолепная технология нуждалась в доработке. Затем в 2007 г. в Стэнфорде я участвовал в конкурсе роботов-автомобилей DARPA Urban Challenge.

В Викторвилле я впервые видел столько беспилотных автомобилей одновременно. Всю команду Стэнфорда просто заворожили машины, мчавшиеся без водителей. Но, что интересно, через пять минут это зрелище стало настолько привычным, что мы отвернулись и принялись беседовать.

М. Ф.: Расскажите, когда начнет использоваться технология компании Drive.ai?

Э. Ы.: Этот сервис уже работает в Техасе. Можно даже посмотреть, что происходит в текущий момент. Вот я вижу, что кто-то использует одно такси. Мне нравится, что это становится обыденностью. Человек вызывает такси, чтобы поехать по своим делам, и к нему приезжает автомобиль без водителя.

М. Ф.: Совпадают ли темпы прогресса в сфере беспилотных автомобилей с вашими ожиданиями?

Э. Ы.: В публичных выступлениях некоторые компании называли нереалистичные сроки появления беспилотных автомобилей на дорогах общего пользования. И хотя я уверен, что эти автомобили изменят нашу транспортную систему в лучшую сторону, мне крайне не нравится ажиотаж, который создают физически невыполнимые обещания.

М. Ф.: Должно ли правительство регулировать использование беспилотных автомобилей и ИИ в целом?

Э. Ы.: Автомобильная промышленность всегда жестко регулировалась для обеспечения безопасности, но массовое использо-

вание беспилотных автомобилей потребует дополнительного регулирования. Страны, которые тщательно пропишут правила пользования, быстрее смогут предоставить доступ к услугам здравоохранения, транспорта или образования на базе ИИ. Страны, где этому аспекту не уделяется достаточно внимания, рискуют остаться позади.

При этом регулировать нужно отдельные сферы применения ИИ, а не ИИ в целом. Это будет способствовать росту самих сфер и позволит искать оптимальные ИИ-решения для каждого конкретного случая. Я считаю, что любой технологический прорыв нуждается в регулировании со стороны правительства. Например, в Сингапуре каждому пациенту присваивается уникальный идентификатор, и все медицинские записи сведены в одну общую систему. Разумеется, в большой стране организовать подобное сложнее, но такие вещи сильно влияют на работу системы здравоохранения.

М. Ф.: С вашей точки зрения, правительство должно участвовать не только в сферах применения ИИ?

Э. Ы.: Я считаю, что правительство должно способствовать развитию ИИ и использовать эту технологию для оптимизации своей работы. Например, для распределения государственных служащих, управления лесными ресурсами, выбора экономической политики, предотвращения налогового мошенничества.

Общественно-государственное партнерство будет способствовать росту промышленности. В Калифорнии компаниям, которые занимаются этими беспилотными автомобилями, не разрешается делать ряд вещей, в итоге им приходится работать в других местах.

М. Ф.: Вы считаете возможным массовое сокращение рабочих мест?

Э. Ы.: Это серьезная этическая проблема. Технологии на базе ИИ уже обогатили некоторые сегменты общества, но мы хотим

не просто способствовать увеличению количества благ, но и их справедливому распределению. Это одна из причин, почему я активно занимаюсь онлайн-образованием и предлагаю доступное переобучение. Шумиха, которая в последнее время поднялась по поводу роботов-убийц, отвлекает от гораздо более сложных и важных социальных проблем, таких как потенциальная безработица.

М. Ф.: Можно ли решить эту проблему введением универсального базового дохода?

Э. Ы.: Эта идея мне не нравится. Куда лучше условный базовый доход: безработным целесообразнее платить пособие, если они учатся. Это увеличивает вероятность того, что человек получит навыки, найдет работу и будет платить налоги, которые пойдут на выплату условного базового дохода.

Продолжение обучения улучшит качество жизни людей. В отчетах экономистов встречаются пугающие утверждения. Например, что за следующие 20 лет из-за автоматизации исчезнет 50 % рабочих мест. При этом остальные 50 % никуда не денутся. Более того, в ряде отраслей наблюдается дефицит кадров. В Соединенных Штатах не хватает медработников, учителей и даже специалистов по ветряным турбинам. Мы уже сейчас не в состоянии найти достаточное количество специалистов и заполнить свободные рабочие места.

Мир, в котором человек мог всю жизнь делать карьеру в одной области, уходит в прошлое. Представители поколения Y легко меняют рабочие места, переходя из одной компании в другую. Уже сейчас глубоким обучением занимаются люди, которые специализируются не в computer science, а в других отраслях, например физике, астрономии или математике.

М. Ф.: Что лучше изучать тем, кто хочет сделать карьеру, связанную с ИИ или с глубоким обучением?

Э. Ы.: Я считаю, что основной упор нужно делать на computer science, машинное обучение и глубокое обучение. Знание когнитивной психологии или физики может пригодиться, но это не самый эффективный путь для карьеры, связанной с ИИ. Сейчас в интернете множество материалов, позволяющих самостоятельно шаг за шагом знакомиться с этой темой. Большое начинается с малого, и я уверен, что при должном упорстве стать специалистом в этой области сможет кто угодно. Тем, кто хочет поменять свою сферу деятельности, я рекомендую подумать, какие возможности могут появиться на вашей текущей работе благодаря ИИ. Намного лучше приобрести дополнительные навыки и стать уникальным специалистом в своей области, а не соревноваться с выпускниками колледжей, которые изучали только ИИ.

М. Ф.: Какие еще опасности, на ваш взгляд, несет ИИ помимо влияния на рынок труда?

Э. Ы.: Мне нравится сравнивать ИИ с электричеством. Любое мощное технологическое достижение можно использовать как во благо, так и во зло. И только от людей зависит, во что все это выльется. Например, я считаю, что нужно обратить пристальное внимание на проблему предвзятости. Ведь, обучаясь на собранных людьми данных, ИИ наследует и различные предрассудки. Я рад, что сейчас об этом говорят и ведется активная работа, направленная на уменьшение предвзятости ИИ.

М. Ф.: Устранить предвзятость в программном обеспечении проще, чем в обществе.

Э. Ы.: Да, в программе можно убрать, например, дискриминацию по половому признаку. При этом у нас нет эффективных способов уменьшения предвзятости у людей. Думаю, системы ИИ в этом нас обгонят.

М. Ф.: Существует ли опасность того, что сильный ИИ в какой-то момент вырвется из-под нашего контроля и превратится в реальную угрозу для человечества?

Э. Ы.: Беспокоиться о роботах-убийцах — это все равно что переживать по поводу перенаселения Марса, колонизация которого не начиналась. Как можно искать решение проблемы, которая еще не появилась?

М. Ф.: То есть вы не верите в сценарий, при котором благодаря рекурсивному улучшению сильный ИИ станет суперинтеллектом?

Э. Ы.: Шумиха по поводу суперинтеллекта и экспоненциального роста по большей части основывается на крайне наивных и упрощенных экстраполяциях. Верить в риск почти мгновенного появления суперинтеллекта из ниоткуда — все равно что верить в возможность перенаселения Марса за одну ночь.

М. Ф.: А что вы думаете о соперничестве с Китаем? Говорят, что Китай имеет преимущества, например, в виде доступа к большему количеству данных и меньшей озабоченности вопросами конфиденциальности.

Э. Ы.: В борьбе за электричество есть выигравшие? Просто какие-то страны электрифицированы больше, какие-то меньше. Гонка в сфере ИИ не так активна, как это показывает популярная пресса. ИИ дает удивительные возможности, и пусть каждая страна самостоятельно решает, каким образом их использовать.

М. Ф.: Но ведь ИИ может применяться для создания автономного оружия. В ООН уже ведутся дебаты по поводу запрета таких видов вооружений.

Э. Ы.: Еще военные используют в своих целях такие вещи, как двигатель внутреннего сгорания, электричество и интегральные схемы. ИИ в этом плане ничем не отличается от остальных технологий.

М. Ф.: Вы очень оптимистично смотрите на происходящее. Я полагаю, что, с вашей точки зрения, по мере развития ИИ польза от этой технологии будет перевешивать опасности, которые она несет?

Э. Ы.: Да. Мне повезло быть в числе пионеров, работавших над продуктами на базе ИИ. И я убедился, насколько большую пользу приносят такие вещи, как распознавание речи, усовершенствованный поиск в интернете и оптимизированные логистические системы.

Я допускаю, что мои взгляды наивны. Мир усложняется и становится не таким, каким я бы хотел его видеть. Честно говоря, я скучаю по временам, когда можно было принимать речи политических и бизнес-лидеров за чистую монету.

Я бы хотел вернуть себе уверенность в том, что компании и лидеры ведут себя этично и имеют в виду именно то, что говорят. Я хочу, чтобы демократия работала лучше, мир был справедливее, люди стали внимательнее друг к другу. Чтобы каждый имел доступ к образованию и был готов усердно работать, расширяя знания и навыки. К сожалению, в реальности все не так, как хотелось бы.

Любой технологический прорыв — это шанс что-то поменять. Я хотел бы, чтобы моя команда вместе со многими другими людьми попыталась сделать мир лучше. Наверное, я просто мечтатель, но упрямый.

М. Ф.: Это прекрасное желание. Вопрос в том, как направить общество в сторону такого будущего. Как вы думаете, есть ли шанс, что мы сделаем правильный выбор?

Э. Ы.: Вряд ли это получится напрямую, но я надеюсь, что в мире достаточно честных, этичных, благонамеренных людей, и шанс есть.

> *Всерьез бояться захвата власти роботами — значит, признать бессилие людей. Нельзя забывать, что именно мы проектируем эти системы, знаем, как они функционируют, и можем их отключить.*

Рана эль Калиуби

ГЕНЕРАЛЬНЫЙ ДИРЕКТОР И СОУЧРЕДИТЕЛЬ КОМПАНИИ AFFECTIVA

Рана эль Калиуби активно участвует в международных форумах, посвященных этическим вопросам и обеспечению положительного влияния ИИ на общество. Была выбрана молодым глобальным лидером на Всемирном экономическом форуме 2017 г. Степени бакалавра и магистра получила в Американском университете в Каире, а докторскую степень — в компьютерной лаборатории Кембриджского университета. В Media Lab MIT Рана разрабатывала технологию помощи детям с расстройствами аутистического спектра. Вошла в список 40 женщин — основателей компаний, которые достигли успеха в 2016 г., составленный интернет-изданием TechCrunch.

Мартин Форд: Как появился ваш интерес к ИИ?

Рана эль Калиуби: Я родилась в Каире, большую часть детства провела в Кувейте. Рано начала экспериментировать с компьютерами. Папа приносил домой старые машины Atari и разбирал их со мной. Затем я получила степень бакалавра computer science в Американском университете в Каире. Меня заинтересовало влияние технологий на способы связи между людьми. Ведь сейчас большая часть общения происходит с помощью технологий.

Затем я получила стипендию для обучения на факультете computer science в Кембриджском университете, что было довольно необычно для молодой египтянки и мусульманки. Это было в 2000 г., когда никто даже не представлял, что может появиться такая штука, как смартфон, но тема взаимодействия человека с компьютером казалась мне актуальной.

В какой-то момент я поняла, что такая личная вещь, как ноутбук, с которым я ежедневно провожу много часов, знает, что я делаю и где нахожусь, но не замечает моего эмоционального состояния.

Он напоминал помощника в Microsoft Office. Забавный анимированный персонаж в форме скрепки появлялся в самый неподходящий момент и спрашивал: «Вам нужна какая-нибудь помощь?» Я поняла, что эмоциональный интеллект компьютера, оторванный от настроения пользователя, нуждается в изменениях.

Электронные сообщения, которые я писала родным после отъезда, не могли передать моих чувств — были полностью лишены богатства невербальных коммуникаций, которого мы не замечаем в процессе общения лицом к лицу.

М. Ф.: То есть искать технологию, позволяющую машине понимать человеческие эмоции, вас заставил личный опыт? Свою диссертацию вы посвятили именно этой теме?

Р. К.: Да. Меня удивляло, почему мы встраиваем в технологии обычный интеллект, игнорируя эмоциональный. Именно с этого вопроса началась моя работа над докторской диссертацией. Затем во время одного из выступлений в Кембридже, когда я рассказывала, что хотела бы научить компьютер обращать внимание на мимику человека, какой-то аспирант спросил: «Вы знаете, что люди с расстройствами аутистического спектра тоже с трудом понимают выражения лиц и невербальное поведение?» С этого началось мое тесное сотрудничество с Кембриджским центром исследования аутизма.

Я позаимствовала собранные ими данные и начала обучать созданные мной алгоритмы распознавать эмоции. Результаты были многообещающими. Кроме выражений радости и грусти я смогла рассмотреть и такие нюансы, как растерянность, интерес, беспокойство и скука.

Когда этот инструмент смог послужить учебным пособием для людей, страдающих аутизмом, я утвердилась в желании продолжить работу.

После защиты диссертации я встретилась с профессором MIT Розалинд Пикард, которая написала книгу *Affective Computing* («Аффективные вычисления»). Именно с ней мы позже создали компанию Affectiva. Уже в 1998 г. Розалинд утверждала, что технологии должны уметь распознавать человеческие эмоции и реагировать на них.

Тогда наша беседа завершилась приглашением присоединиться к ее лаборатории Media Lab. Это был проект Национального научного фонда, в рамках которого моя технология чтения эмоций с интегрированной камерой применялась для работы с детьми с расстройствами аутистического спектра.

М. Ф.: Вы говорите об «эмоциональном слуховом аппарате» для детей с аутизмом. Это уже продукт или все осталось на уровне концепции?

Р. К.: В 2006–2009 гг. мы сотрудничали со школой для особых детей в Провиденсе, штат Род-Айленд. Мы предлагали детям протестировать нашу технологию и совершенствовали ее в зависимости от отзывов. Наконец, дети научились устанавливать зрительный контакт, а не просто смотреть на лица собеседников.

Они надевали очки с прикрепленной камерой, которая сначала транслировала пол или потолок, потому что на лицо они старались не смотреть. Но постепенно мы смогли наладить обратную связь в режиме реального времени, помогающую установлению зрительного контакта. После этого мы начали объяснять детям, какие эмоции проявляют люди. Это была очень интересная работа.

Напомню, что Media Lab — это уникальный отдел MIT, имеющий настолько прочные связи с промышленностью, что около 80 % финансирования поступает к нему от компаний из списка Fortune Global 500. Два раза в год у нас проходит так называемая Неделя спонсоров, во время которой мы рассказываем о проделанной работе. Если не будет результатов, проект просто перестанут финансировать.

После демонстрации прототипа у компании PepsiCo возник вопрос, нельзя ли применять наши наработки для проверки эффективности рекламы. Procter & Gamble хотела с их помощью определить отношение потребителей к их продукции. Toyota задумалась о возможностях мониторинга состояния водителей. A Bank of America — об оптимизации банковской деятельности. Мы решили привлечь еще несколько научных сотрудников для развития идей, которые интересовали наших спонсоров. Вскоре стало понятно, что речь идет уже не об исследованиях, а о коммерческой возможности.

Отход от научной работы меня немного пугал. Было понятно, что в существующих условиях о масштабном применении наших

прототипов речи не шло, но появилась возможность выводить продукты на рынок и применять новые способы общения.

М. Ф.: Похоже, компания Affectiva имеет высокую клиентоориентированность.

Р. К.: Вы абсолютно правы. Мы с Розалинд ощутили себя родоначальниками нового течения и идейными лидерами, ответственными за этическую составляющую того, что мы делаем.

М. Ф.: Над чем сейчас работает компания Affectiva и какие у вас планы на будущее?

Р. К.: Миссия компании заключается в гуманизации технологий. Ведь технологии постепенно проникают в каждый аспект жизни. Интерфейсы все больше способны к диалогу, а устройства делаются восприимчивыми и социальными. У людей возникают своеобразные отношения с автомобилями, телефонами и виртуальными помощниками, такими как Alexa от Amazon или Siri от Apple.

Создатели этих устройств в настоящее время сосредоточены на таком аспекте, как когнитивный интеллект, и забывают об интеллекте эмоциональном. Но ведь успех человека в профессиональной и личной жизни определяется не только уровнем его IQ. Он зависит от того, насколько человек понимает состояние окружающих, может адаптировать свое поведение и влиять на других.

Все эти вещи требуют развитого эмоционального интеллекта. Если технология взаимодействует с людьми на регулярной основе, помогает человеку лучше спать, питаться, больше заниматься спортом, продуктивнее работать или развивать социальные навыки, она должна учитывать его психическое состояние. Я уверена, что постепенно такие интерфейсы станут повсеместными.

М. Ф.: Я слышал, что вы разрабатываете способ мониторинга концентрации водителей на дороге?

Р. К.: Да, но проблема в том, что существует множество ситуаций, которые нужно предусмотреть. Мы стараемся выбирать те из них, которые упираются в этические вопросы, и, конечно, ориентируемся на востребованность своих предложений на рынке.

Компания Affectiva начала свою деятельность в 2009 г. с тестирования рекламы, и, как я уже упоминала, сегодня мы обслуживаем четверть компаний из списка Fortune Global 500, помогая им понять, какой эмоциональный отклик вызывает у потребителей их реклама. До появления наших технологий единственным способом понять, достигнут ли желаемый эффект, были опросы. Тем, кто посмотрел рекламу, предлагали ответить, понравилась ли она, была ли смешной, вызвала ли желание купить продукт. Собранные таким способом данные нельзя считать достоверными.

Наша технология с согласия зрителей анализирует выражения их лиц во время просмотра рекламы. Постепенно накапливается массив данных, который позволяет сделать объективный вывод о том, какую именно эмоциональную реакцию вызывает конкретная реклама. Эти данные можно сопоставить, например, с желанием совершить покупку, сведениями о продажах и виральностью.

Наш продукт может отслеживать все ключевые показатели эффективности и связывать эмоциональный отклик с фактическим поведением потребителей. Им пользуются в 87 странах, от США до Китая и Индии. Он дает довольно надежные результаты и позволяет собирать данные со всего мира. Я бы сказала, что таких данных нет даже у компаний Facebook и Google. Ведь мы рассматриваем не фотографии в профилях, а живую реакцию.

М. Ф.: Вы анализируете в основном выражение лица или учитываете и другие вещи, например голос?

Р. К.: Года полтора назад мы решили рассмотреть не только мимику, но и другие факторы, позволяющие судить о реакции.

Известно, что при считывании состояния собеседника примерно 55 % данных человек берет от выражения его лица и жестов, около 38 % — от тона его голоса, скорости и энергичности речи и только 7 % извлекается из значения использованных слов!

Получается, что анализ тональности текстов путем выявления эмоционально окрашенной лексики задействует всего 7 % передаваемой в процессе общения информации. Мне нравится думать, что мы пытаемся охватить остальные 93 %, приходящиеся на невербальное общение.

Я создала группу, которая рассматривает просодические паралингвистические особенности общения, к которым относятся тон голоса и речевые события, такие как употребление междометий или смех. Мы используем так называемый мультимодальный подход.

М. Ф.: Зависят ли индикаторы эмоций, которые вы рассматриваете, от языка и культуры?

Р. К.: Такие базовые вещи, как выражение лица или даже тон голоса, универсальны. Улыбка везде остается улыбкой. Но существуют и дополнительные культурные нормы выражения эмоций. Люди могут усиливать выражение реакции, ослаблять или маскировать. Например, жители Азии менее склонны к проявлению негативных эмоций. Поэтому там можно говорить о такой вещи, как социальная улыбка, которая является просто проявлением вежливости, а не выражением радости.

Мы собрали уже такой большой массив универсальных данных, что смогли построить модели региональных норм и даже норм для отдельных стран, например для Китая. Именно это позволяет успешно проводить мониторинг эмоциональных состояний по всему миру.

М. Ф.: Тогда можно предположить, что вы работаете и над приложениями, направленными на увеличение безопасности, которые, к примеру, следят за степенью концентрации водителей и операторов опасного оборудования?

Р. К.: Да, с прошлого года автомобильная промышленность проявляет к нам интерес. Это здорово, потому что не только открывает перед компанией Affectiva большие рыночные возможности, но и позволяет поработать над двумя интересными проблемами.

Сейчас уже появились полуавтономные транспортные средства, такие как Tesla, но проблема безопасности по-прежнему стоит очень остро. Программное обеспечение от Affectiva позволяет проверять такие вещи, как сонливость, рассеянное внимание, усталость и даже опьянение. В последнем случае водитель получает предупреждение или программа вмешивается в процесс управления. Вмешательство может принимать разные формы, от смены музыки до струйки холодного воздуха, от затягивания ремня безопасности до высказывания: «В настоящий момент я могу обеспечить более безопасное вождение и поэтому беру на себя контроль».

Вторая задача, над которой мы работаем, связана с обслуживанием пассажиров. Ведь в условиях широкого применения полностью автономных транспортных средств важно, чтобы автомобиль мог сам определять состояние пассажиров: их количество, какие отношения их связывают, беседуют ли они друг с другом, есть ли в машине ребенок. Зная все эти вещи, он сможет персонализировать пользовательский опыт.

Автоматика способна рекомендовать маршрут или какие-то продукты. Фактически автомобильные компании, особенно высшего класса, такие как BMW или Porsche, получают новую бизнес-модель. До сих пор речь шла только об опыте вождения. Современный транспорт — очень перспективный рынок, и мы прикладыва-

ем все усилия для создания новых решений, которые потребуются в этой отрасли.

М. Ф.: Могут ли ваши технологии помочь в таких областях, как психологическое консультирование?

Р. К.: Я считаю, что у наших технологий больше перспективы в сфере здравоохранения. Например, известно, что существуют лицевые и голосовые биомаркеры депрессии и признаки, по которым можно предсказать суицидальные намерения. Если учесть, сколько времени люди проводят перед различными устройствами, понятно, что мы можем собрать объективные данные.

В настоящее время диагностика сводится к тому, что человека просят оценить по шкале от 1 до 10, насколько он подавлен. Мы же даем возможность построить базовую модель для оценки психического здоровья. Если состояние человека начнет отклоняться от базовых показателей, система может подать сигнал самому человеку, членам его семьи или даже медицинскому работнику.

Эти метрики применимы и для анализа эффективности различных методов лечения. Например, человек пробует методы когнитивно-поведенческой терапии или начинает принимать препараты, а система количественно определяет эффективность применяемого лечения.

М. Ф.: Насколько этичными вы считаете свои разработки? Представьте, что ваша система используется во время переговоров для тайного наблюдения за оппонентами. Получаемая при этом информация создает несправедливое преимущество.

Р. К.: Вопрос о границах мы с Розалинд и нашим первым сотрудником обсуждали перед началом тестирования. Эмоции — это личные данные, поэтому мы договорились, что действовать можно только в тех случаях, когда люди прямо соглашаются ими поделиться.

В 2011 г. мы столкнулись с недостатком средств, и можно было получить финансирование от охранного агентства, которое имело венчурное подразделение. Они хотели использовать наши технологии для наблюдения и обеспечения безопасности. По большей части люди осведомлены, что в аэропортах за ними следят, но все равно мы отклонили это предложение, так как чувствовали, что оно противоречит нашему основополагающему принципу.

Один и тот же инструмент может как принести пользу, так и создать эффект присутствия Большого Брата. Но если бы сотрудники какой-то фирмы согласились на установку наблюдения, работодатель смог бы оценить уровень комфорта на рабочем месте.

Если руководитель проводит онлайн-презентацию, то с помощью технологий может дистанционно оценить уровень мотивации сотрудников к работе над новой задачей. Но оценка внимания при прослушивании презентации повлияет на личные границы человека.

Одна из версий нашей технологии умеет отслеживать ход встреч и в конце давать всем участникам обратную связь. То есть человек может услышать, к примеру: «Вы 30 минут говорили о пустяках и проявили враждебность к коллеге N, нужно быть более вдумчивым и чутким». Легко представить, как с помощью этой технологии коуч учит персонал лучше вести переговоры. Но также легко воспользоваться ею, чтобы навредить чужой карьере.

Мы предпочитаем давать людям обратную связь для развития социальных навыков и эмоционального интеллекта.

М. Ф.: В своей работе вы активно используете глубокое обучение. Но говорят, что прогресс в этой области замедлился и даже может совсем остановиться. Вы согласны с необходимостью искать другие подходы?

Р. К.: Работая над докторской диссертацией, для количественной оценки и построения классификаторов я использовала динами-

ческие байесовские сети. Пару лет назад на глубоком обучении начала базироваться вся наша инфраструктура, что принесло многочисленные плоды.

Я бы сказал, что максимальные результаты, которые может дать глубокое обучение, еще не достигнуты. Увеличение количества данных в сочетании с глубокими нейронными сетями позволяет проводить точный и достоверный анализ в самых разных ситуациях.

Но при этом, конечно, всех существующих потребностей глубокое обучение не удовлетворяет, и требуются другие методы и подходы.

М. Ф.: Что, с вашей точки зрения, мешает созданию сильного ИИ? Ожидаете ли вы увидеть его при жизни?

Р. К.: До создания сильного ИИ еще очень далеко, ведь все существующие примеры ИИ имеют достаточно узкое применение. Современные ИИ-приложения прекрасно справляются с конкретными задачами, для решения которых они создавались, но при этом все они требуют начальной загрузки, определенных допущений и помощи людей. Самая лучшая система обработки естественного языка не может пройти тест, с которым справляются школьники.

М. Ф.: Вы пытаетесь научить машины понимать эмоции, а должны ли машины их демонстрировать?

Р. К.: Над машинами, проявляющими эмоции, мы тоже работаем. Компания Affectiva разработала платформу для восприятия эмоций, которую многие наши партнеры применяют для управления поведением машин. В качестве входных данных берутся человеческие метрики. Ответы робота зависят от действий человека. По такому принципу работает, к примеру, виртуальный помощник Alexa от Amazon.

Когда, например, Alexa раз за разом ошибается, делая заказ по вашей просьбе, это раздражает. Но вместо того чтобы просто игнорировать этот момент, помощник может сказать: «Прошу

прощения. Я понимаю, что ошибаюсь. Разрешите мне попробовать еще раз». То есть устройство учитывает раздражение пользователя и соответствующим образом корректирует свои действия. Робота можно запрограммировать на движения, показывающие, что он сожалеет о сделанной ошибке. Разумеется, все это не означает, что устройства на самом деле испытывают эмоции. Но пользователям это и не нужно.

М. Ф.: Как появление ИИ может повлиять на экономическую ситуацию и рынок труда? Стоит ли волноваться по поводу грядущей массовой безработицы?

Р. К.: Я предпочитаю думать о партнерском взаимодействии человека и технологий. Конечно, часть рабочих мест исчезнет, но в истории человечества такое случалось часто. На смену им приходили новые рабочие места и новые возможности. Подозреваю, что могут появиться специальности, которых мы сейчас не можем даже представить. Не думаю, что мы окажемся в мире, где всю работу возьмут на себя роботы, а люди будут просто отдыхать. Мое детство на Ближнем Востоке пришлось на время первой войны в Персидском заливе, я понимаю, что в мире множество проблем, и не верю, что появится машина, которая все решит. Для людей всегда останется работа.

М. Ф.: Мне кажется, что благодаря вашим разработкам можно автоматизировать выполнение работы, связанной с общением и наличием эмоционального интеллекта.

Р. К.: Да, вы правы. Например, в Affectiva разрабатываются виртуальные медсестры и роботы, помогающие неизлечимо больным пациентам. Я сомневаюсь, что они смогут заменить человека, но, скорее всего, набор рабочих обязанностей медсестры со временем изменится.

Когда робот при обнаружении невыполнимой задачи привлекает медсестру-человека, растет количество пациентов, которые

одновременно могут получать помощь. Аналогичная ситуация с учителями. Такие разработки полезны в ситуации нехватки персонала.

Мне кажется, что профессия водителя грузовика тоже исчезнет лет через десять. Достаточно будет одного оператора, который сидит перед экраном и управляет сотней автомобилей. Но при этом останется и работа для людей-водителей, которые могут при необходимости брать на себя контроль над одной из машин.

М. Ф.: Что вы думаете об экзистенциальных рисках, которые несут обычный и сильный ИИ, и предупреждениях Илона Маска?

Р. К.: В интернете можно найти документальный фильм «*Доверяете ли вы этому компьютеру?*», создание которого частично финансировал Илон Маск. Там я подробно высказалась на эту тему.

М. Ф.: Да, я смотрел его. В нем участвуют и другие ученые, у которых я брал интервью для этой книги.

Р. К.: Я убеждена, что люди приносят куда больше проблем, чем ИИ. Всерьез бояться захвата власти роботами — значит признать бессилие людей. Нельзя забывать, что именно мы проектируем эти системы, знаем, как они функционируют, и можем их отключить. Поэтому я не разделяю подобных страхов. С ИИ-системами связаны более актуальные проблемы. Например, как избежать закрепления существующих предрассудков?

М. Ф.: Вы считаете предвзятость одной из наиболее актуальных проблем?

Р. К.: Да. Технология развивается очень быстро, но мы не до конца понимаем, чему именно мы учим нейронные сети. Я боюсь, что мы просто закрепляем в них все существующие в обществе предубеждения.

М. Ф.: Получается, что предвзятость закладывается на уровне данных.

Р. К.: Важны и сами данные, и то, как мы их применяем. Поэтому в Affectiva мы стараемся убедиться в том, что для обучения алгоритмов берутся данные, репрезентативные для всех этнических групп, сбалансированные с гендерной и возрастной точки зрения. Это непрекращающаяся работа, ведь для защиты от смещений всегда можно сделать больше.

М. Ф.: Бороться с необъективностью людей очень сложно, и, наверное, проще убирать ее проявления из алгоритмов, чтобы в будущем сами алгоритмы помогали уменьшить количество предвзятости и дискриминации в мире.

Р. К.: Именно так. Я даже могу привести отличный пример. Компания HireVue использует нашу технологию для процедуры найма. Кандидаты присылают видеоинтервью, и система сортирует их, применяя комбинацию наших алгоритмов и классификаторов обработки естественного языка. При этом учитываются не только ответы на вопросы, но и невербальные сигналы. Алгоритм не обращает внимания на половую и этническую принадлежность.

Результаты применения цифровой платформы для проведения собеседований в компании Unilever показали не только сокращение времени найма на 90 %, но и увеличение многообразия нанятых сотрудников на 16 %. Мне кажется, это очень впечатляющий результат.

М. Ф.: Нужно ли регулирование в сфере ИИ? Вы говорили, что у компании Affectiva очень высокие этические стандарты, но кто-то может разработать аналогичные технологии и продать их, например, для тайного наблюдения за людьми.

Р. К.: Я большой сторонник регулирования. Компания Affectiva входит в консорциум Partnership on AI и в рабочую группу FATE (Fair, Accountable, Transparent, Equitable). Это аббревиатура оз-

начает: справедливый, подконтрольный, прозрачный и беспристрастный ИИ.

Задача этих групп заключается в разработке руководящих принципов для ИИ, соответствующих тем, которые разрабатывает Управление по санитарному надзору за качеством пищевых продуктов и медикаментов. Кроме того, Affectiva публикует лучшие практики и рекомендации для отрасли. Как лидеры, мы обязаны быть сторонниками регулирования и предпринимать активные действия, а не просто ждать появления нужного закона.

Мы принимаем участие во Всемирном экономическом форуме, в рамках которого работает совет по робототехнике и ИИ. Меня поразило, насколько по-разному в разных странах воспринимается эта сфера. Например, в Китае никого не заботят этические вопросы, что отражается на регулировании ИИ.

М. Ф.: Чтобы закончить на позитивной ноте, скажите, вы оптимист? Вы верите, что в итоге эти технологии принесут пользу человечеству?

Р. К.: Да, я могу назвать себя оптимисткой. Я считаю, что технологии нейтральны. Важно то, как ими воспользуются. Мне бы хотелось, чтобы вся отрасль следовала по стопам моей команды.

> *Согласно моему сценарию, в кровеносную систему можно будет вводить медицинских нанороботов. <…> А нанороботы в числе прочего проникнут к нам в мозг и позволят напрямую подключаться к виртуальной и дополненной реальности.*

Рэймонд Курцвейл

ТЕХНИЧЕСКИЙ ДИРЕКТОР КОМПАНИИ GOOGLE, СОУЧРЕДИТЕЛЬ УНИВЕРСИТЕТА СИНГУЛЯРНОСТИ

Рэймонд Курцвейл — один из ведущих мировых изобретателей, мыслителей и футурологов, обладатель 21 почетной докторской степени и почетных званий от трех президентов США.

Специальность инженера получил в MIT под руководством Марвина Минского. Изобрел первый сканер с ПЗС-матрицей, работал над развитием первой системы оптического распознавания символов, синтезатора, преобразующего текст в речь, первого музыкального синтезатора, способного воссоздавать звуки оркестровых инструментов, и первого коммерческого приложения для распознавания речи с большим словарным запасом.

Получил премию «Грэмми» за выдающиеся достижения в области музыкальных технологий, Национальную медаль США в области технологий, введен в Национальный зал славы изобретателей. Написал пять книг, ставших бестселлерами, в том числе «Син-

гулярность уже близка» (2005 г.) и How to create a mind: the secret of human thought revealed («Как создать разум», 2012 г.). Первый роман Курцвейла Danielle, Chronicles of a Superheroine («Даниэль. Хроники супергероини») вышел в начале 2019 г., а к концу года ожидается выход книги The Singularity is Nearer («Сигнулярность еще ближе»).

Известен утверждениями об экспоненциальном прогрессе в технологиях, которые сформулировал как закон ускоряющейся отдачи.

Мартин Форд: Как возник ваш интерес к ИИ?

Рэймонд Курцвейл: Я начал заниматься ИИ в 1962 г., всего через шесть лет после того, как Марвин Минский и Джон Маккарти придумали этот термин на Дартмутском семинаре в Ганновере, штат Нью-Гэмпшир.

К тому моменту появились два враждующих лагеря: сторонники символьной системы и коннекционисты. Доминировали первые, а их лидером считался Марвин Минский. Коннекционистов считали выскочками. Одним из них был Фрэнк Розенблатт из Корнеллского университета — создатель первой популяризованной нейронной сети, названной персептроном. Я написал им, и получил приглашения в гости. Сначала я навестил Минского, с которым мы провели целый день. Так начались наши отношения, которые продлились 55 лет. Мы обсуждали ИИ, который в то время рассматривался как нечто непонятное и не представляющее особого интереса.

Затем я встретился с Розенблаттом. Его персептрон представлял собой устройство с камерой. Я принес с собой распечатки писем, и устройство прекрасно распознавало весь текст, который был набран шрифтом Courier 10. Остальные шрифты не распознавались,

но Розенблатт сказал: «Не волнуйтесь, я могу взять выход персептрона, передать его на вход второму персептрону, а полученный выход — на вход третьему слою, и постепенно сеть обучится, обобщит данные и даст результат». Я спросил, пробовал ли он это делать, и узнал, что пока таких экспериментов не проводилось, но они есть в программе исследований.

В 1960-е гг. работа велась медленнее, чем сегодня. И, к сожалению, спустя девять лет Розенблатт погиб, так и не успев опробовать эту идею, которая оказалась на удивление пророческой. Все впечатляющие результаты, которые мы имеем сегодня, получены с помощью глубоких нейронных сетей. Так что Розенблатт очень правильно все понимал, хотя на тот момент было неясно, сработает ли такой подход.

В 1969 г. Минский вместе с Сеймуром Пейпертом написал книгу о персептронах. По сути, там доказывалось, что они не могут решать задачу XOR и проблему связности. Обложку книги украшали изображения двух лабиринтов: один — связный, а второй — нет. Доказано, что персептрон не может классифицировать фигуры по признаку связности. Книга привела к тому, что в течение следующих 25 лет на коннекционизм просто не выделяли средств. Впоследствии Минский сожалел об этом и незадолго до своей смерти говорил, что осознал потенциал глубоких нейронных сетей.

М. Ф.: Но ведь он работал над первыми коннекционистскими нейронными сетями еще в 1950-х гг.?

Р. К.: Да, но к 1960 г. он в них разочаровался. Более того, возможности этих сетей стали очевидными только десятилетия спустя, после того как опробовали трехслойные нейронные сети. При большем количестве слоев возникала проблема взрывающегося или исчезающего градиента, то есть ситуации, когда динамический диапазон коэффициентов уменьшался, потому что числа становились слишком большими или слишком маленькими.

Эту проблему решил Джеффри Хинтон с группой математиков. Они предложили после каждого слоя корректировать веса, чтобы они оставались в заданном диапазоне значений. Это дало возможность работать с нейронными сетями, состоящими из ста слоев. Но все равно осталась проблема, которую можно описать фразой: «Жизнь начинается с миллиарда примеров».

Это одна из причин, по которой я работаю в Google. Тут есть доступ к миллиарду примеров изображений собак, кошек и других вещей. Хотя остается и множество вещей, для которых ничего подобного не существует. Скажем, у нас есть многочисленные примеры естественного языка, но их значения не снабжены аннотациями. Да и как это можно сделать для языков, которых мы изначально не понимаем? Для решения некоторых задач можно обойтись без огромного набора маркированных данных. К таким задачам относится, например, игра го. Когда систему, созданную компанией DeepMind, пытались обучать на примере всех онлайн-ходов, нашлось порядка миллиона примеров. Система научилась играть на любительском уровне. Для дальнейшего обучения требовалось еще 999 млн примеров, а где их брать?

М. Ф.: Вы утверждаете, что современное глубокое обучение сильно зависит от наличия маркированных данных, то есть все сводится к обучению с учителем.

Р. К.: Именно так. Это ограничение можно обойти, если получится смоделировать мир, в котором вы работаете. В этом случае появится возможность создавать собственные обучающие данные. Так и поступили в DeepMind. Они просто стали играть все новые и новые партии. Ходы маркировались традиционными методами. В результате без предварительно представленных людьми данных нейронную сеть удалось обучить настолько, что алгоритм AlphaZero смог сто раз подряд победить алгоритм AlphaGo.

Вопрос в том, в каких ситуациях работает такой подход? Например, симулировать можно математические правила: аксиомы

теории чисел не сложнее правил игры го. А процесс вождения намного сложнее. Но в компании Waymo применили комбинацию методов и создали достаточно хороший беспилотник, а затем проехали миллионы миль с водителем, готовым в любой момент взять управление на себя. В результате накопилось достаточно данных для создания точного симулятора. После этого порядка миллиарда миль было пройдено в симуляторе, чтобы сгенерировать обучающие данные для глубокой нейронной сети, предназначенной для улучшения алгоритмов.

Теперь очередь за моделированием биологического и медицинского миров. Подобные симуляторы позволили бы проводить клинические испытания за часы, а не за годы, и генерировать собственные обучающие данные.

Это не единственный подход к проблеме обучающих данных. Например, у людей работает технология переноса обучения (transfer learning), то есть они умеют использовать накопленный опыт для решения задач другого типа. Я придумал собственную модель на базе приблизительного представления о работе неокортекса. Первые идеи появились в 1962 г., и последующие 50 лет я всячески их обдумывал. Вместо одной большой нейронной сети можно обучить на шаблонах множество модулей. В книге «Как создать разум» я описываю неокортекс как 300 млн модулей, последовательно распознающих образ и допускающих определенную степень изменчивости. Они образуют иерархическую структуру. Для такого обучения требуется меньше данных. Машина, как человек, сумеет обобщать информацию и переносить ее из одной предметной области в другую. После выхода книги Ларри Пейдж пригласил меня в Google, чтобы я попробовал применить свои идеи к распознаванию естественного языка.

М. Ф.: На базе этих концепций уже созданы какие-то продукты?

Р. К.: Функция Smart Reply в сервисе Gmail, которая предлагает три базовых варианта ответа на каждое письмо, использует именно

такую иерархическую систему. Недавно мы запустили сервис Talk to Books[1], позволяющий получать ответы из книг. Пользователь задает вопрос, а система за полсекунды читает 100 тысяч книг (примерно 600 млн предложений) и возвращает наиболее подходящий ответ. Все это базируется не на ключевых словах, а на семантическом понимании.

В Google мы сильно продвинулись в работе над естественным языком. Язык создается и понимается неокортексом. Он иерархичен, благодаря чему мы можем делиться друг с другом возникающими в неокортексе идеями. Я думаю, что тест Тьюринга проводится на естественном языке, потому что в понимании речи задействован весь спектр человеческого мышления и интеллекта.

М. Ф.: Ваша конечная цель — создать машину, способную пройти тест Тьюринга?

Р. К.: Мою точку зрения многие не разделяют, но я считаю, что правильно подготовленный тест Тьюринга действительно позволяет проверить, обладает ли сущность интеллектом на уровне человека. Проблема в том, что в кратком документе 1950 г. Тьюринг не описал процедуру тестирования. Я поспорил с Митчем Капором на 20 тысяч долларов, что ИИ пройдет тест Тьюринга к 2029 г. Победитель отдаст выигрыш на благотворительные нужды.

М. Ф.: Но согласитесь, что эффективным тест Тьюринга станет только после снятия ограничения по времени. В течение 15 минут не так уж сложно водить экзаменатора за нос.

Р. К.: Вы совершенно правы. Согласно правилам, которые придумали мы с Митчем, тест занимает несколько часов, и даже этого может оказаться недостаточно. Машина должна убедить экзаменатора, что она — человек. Такую иллюзию вполне можно создать с помощью несложных уловок.

[1] https://books.google.com/talktobooks/

М. Ф.: Тест покажет, что машина обладает интеллектом, не найдя в ней сходство с человеком.

Р. К.: Киты и осьминоги проявляют интеллектуальное поведение, но никогда не смогут пройти тест Тьюринга. Китаец, который не говорит по-английски, тоже не пройдет. Невозможность пройти тест не указывает на отсутствие развитого интеллекта, но чтобы его пройти, нужно обладать развитым интеллектом.

М. Ф.: Вы верите в то, что комбинация глубокого обучения и вашего иерархического подхода позволяет двигаться в сторону сильного ИИ или для этого нужен какой-то сдвиг парадигмы?

Р. К.: Я думаю, что люди используют именно иерархический подход. Каждый из модулей способен обучаться, и в книге я подчеркиваю, что в человеческом мозге происходит не глубокое обучение, а какой-то эквивалент марковского процесса. При этом глубокое обучение результативно. В системах Google мы с его помощью создаем векторы, представляющие собой шаблоны для каждого модуля иерархии, а затем начинается сама иерархия, которая в парадигму глубокого обучения уже не вписывается. Я думаю, что для сильного ИИ этого достаточно. На мой взгляд, иерархический подход отражает процессы, которые протекают в мозге человека, и проекты по реверсивному воспроизведению мозга это подтверждают.

Существуют доказательства того, что мозг использует систему, основанную на правилах, а не ту, что предлагают коннекционисты. Именно поэтому люди способны четко отличать вещи друг от друга и логически мыслить. При этом коннекционистская система в определенных ситуациях настолько уверена в своих суждениях, что действует как система, основанная на правилах, и при этом справляется с исключениями и нюансами.

Обратное при этом неверно. Система, основанная на правилах, не может эмулировать коннекционистскую систему. Проект *Cyc* (чи-

тается «сайк») Дугласа Лената очень впечатляет, но, на мой взгляд, он страдает от ограничений системы, основанной на правилах. В какой-то момент неизбежно достигается предел сложности. То есть правила становятся настолько сложными, что при попытке исправить одну вещь ломаются три другие.

М. Ф.: *Cyc* — это проект, в котором люди вручную вводят логические правила?

Р. К.: Да. Точных цифр я не знаю, но количество правил там огромно. В одном из режимов можно распечатывать рассуждения, обосновывающие поведение. Обычно они занимают несколько страниц, и уследить за их ходом сложно. Это хороший проект, но человеческий интеллект формируется как-то по-другому. Вместо каскадов правил люди используют самоорганизующуюся иерархию.

Преимуществом коннекционистского подхода я считаю и то, что он прозрачен. Можно посмотреть на модули в иерархии и увидеть, на какие решения влияет каждый из них. А состоящие из 100 слоев нейронные сети действуют как большой «черный ящик». Понять ход происходящих внутри рассуждений крайне сложно, хотя некоторые пытались.

М. Ф.: В человеческом мозге с рождения присутствуют определенные структуры, например, позволяющие новорожденным распознавать лица.

Р. К.: У нас есть генераторы функций. Веретенообразная извилина умеет вычислять соотношения: длину носа или расстояние между глазами. Существует примерно дюжина простых функций, которые мы можем сгенерировать на основе изображения лица, а затем распознать то же самое лицо на новом изображении, даже если какие-то детали изменились. Другие функции аналогичным образом работают с аудиоинформацией, то есть вычисляют определенные отношения и распознают частичные обертоны. Затем

эти функции поступают в коннекционистскую иерархическую систему.

М. Ф.: Расскажите, пожалуйста, о перспективах создания сильного ИИ или интеллекта уровня человека.

Р. К.: Это синонимы, но термин «сильный ИИ» мне не нравится. Разработки ИИ с самого начала ставили целью достижение человеческого уровня интеллекта. Но постепенно начали формироваться отдельные области, исчезла концентрация на сильном ИИ. Но я считаю, что, решая отдельные задачи, мы постепенно дойдем и до него.

Не стоит забывать и о разнице в уровне навыков при выполнении одной и той же задачи. Насколько хорошо люди играют в го? Как только компьютер достигает хотя бы нижнего уровня способностей человека, он очень быстро может превзойти способности чемпиона.

Компьютеры пока не могут хорошо работать с несколькими цепочками рассуждений, делая выводы из нескольких утверждений и одновременно применяя свои знания об окружающем мире. Например, в тесте на знание языка для третьего класса компьютер не понимал, что у мальчика грязные ботинки из-за ходьбы по лужам, и следы на полу рассердят его мать. ИИ не обладает тем опытом, который делает для нас многие вещи очевидными.

Сейчас в ряде языковых тестов компьютеры демонстрируют среднестатистический уровень понимания взрослого человека. И быстрого прогресса в этой области не будет, потому что сначала нужно решить более фундаментальные проблемы. Но даже достигнутый уровень впечатляет, потому что для понимания языка требуется и высокий интеллект, и умение распознавать переносные смыслы, и иерархическое мышление. Подводя итог, скажу, что да, используя коннекционистские подходы, мы делаем успехи.

Моя рабочая группа нацелена на прохождение теста Тьюринга. Научить компьютеры учитывать выводы и подтексты различных концепций, то есть вести несколько цепочек рассуждений одновременно — это первоочередная задача. Именно тут чат-боты обычно терпят неудачу. Если я скажу, что переживаю из-за школьных оценок дочери, ни один человек не спросит через три хода, есть ли у меня дети. А чат-боты задают такие вопросы. Но если мы научим их понимать все оттенки языка, виртуальные собеседники смогут извлекать все нужные сведения из доступных в интернете книг и документов. У нас уже есть идеи, как реализовать подобные вещи.

М. Ф.: Долгое время вы утверждали, что сильный ИИ будет создан к 2029 г. Вы до сих пор так считаете?

Р. К.: В книге *The Age of Intelligent Machines* («Эпоха мыслящих машин», 1989 г.) я писал о 2029 г. плюс-минус 10 лет. В 1999 г. я опубликовал книгу *The Age of Spiritual Machines* («Эпоха чувствующих машин»), в которой четче обозначил 2029 г. По этому поводу в Стэнфорде прошла конференция экспертов по ИИ. В основном все пришли к мнению, что на это потребуются сотни лет, а примерно четверть присутствовавших считали, что этого не произойдет никогда.

В 2006 г. в Дартмутском колледже состоялась конференция, посвященная 50-летию Дартмутского семинара. Проведенный там опрос показал, что создание сильного ИИ ожидается где-то лет через пятьдесят. В 2018 г. основная масса специалистов называла срок в 20–30 лет. Получается, что я просто более оптимистичен.

М. Ф.: До названной вами даты осталось всего 11 лет. Это совсем немного.

Р. К.: Прогресс идет по экспоненте. Мы уже добились значительных успехов в области беспилотных автомобилей, понимании языка, игре го и во множестве других сфер. Аппаратное и про-

граммное обеспечение развиваются в быстром темпе, причем первое — активнее.

М. Ф.: Мнение, что потребуется более ста лет, означает, что люди считают прогресс линейным?

Р. К.: Во-первых, они мыслят линейно, а во-вторых, подвержены явлению, которое я называю пессимизмом инженера. Они так сосредоточены на одной сложной проблеме, что начинают считать себя полностью ответственными за ее решение и отталкиваются исключительно от собственных темпов работы. Учитывать темпы прогресса в области в целом и то, как идеи взаимодействуют друг с другом, нужно уметь.

Для проекта по определению генома человека 1 % данных собирали семь лет. Скептики тогда называли проект безнадежным, ведь при таких темпах на его реализацию потребовалось бы 700 лет. Я же считал, что работа почти закончена. Потому что каждый год темп ускорялся в два раза, и проект был завершен семь лет спустя.

Независимо от достижений или уровня интеллекта обычный пользователь или специалист могут держаться за линейное мышление.

М. Ф.: Но согласитесь, что речь идет не только об экспоненциальном росте скорости вычислений или объема памяти. Нужны концептуальные прорывы, в результате которых компьютеры получат способность, например, обучаться в реальном времени или на неструктурированных данных, как это делают люди.

Р. К.: Программное обеспечение также экспоненциально прогрессирует, несмотря на не поддающиеся прогнозированию вещи, которые вы упомянули. Существует взаимообогащение, благодаря которому после достижения некоторого уровня производительности появляются идеи для движения дальше.

Научно-экспертный совет администрации Обамы исследовал этот вопрос. Они взяли дюжину классических инженерных и технических проблем и изучили, насколько прогресс связан с оборудованием. В течение предыдущих 10 лет это соотношение, как правило, составляло примерно тысячу к одному. Значит, показатель «цена/производительность» удваивался ежегодно. Ситуация с программным обеспечением варьировалась, но в каждом случае показатель для него был больше, чем для аппаратного обеспечения. Развитие программного обеспечения происходит нелинейно; это геометрическая прогрессия. А для оценки прогресса в целом нужно перемножить показатели прогресса в областях аппаратного и программного обеспечения.

М. Ф.: Следующая дата, которую вы предсказали, это 2045 г. Наступление того, что вы называете сингулярностью. Это означает появление суперинтеллекта?

Р. К.: Лично я считаю, что развитие продолжится по экспоненте. Идея интеллектуального взрыва базируется на том, что в какой-то момент компьютер сможет получить доступ к своему проектированию и, циклически совершенствуя сам себя, сможет быстро достичь суперинтеллекта.

Но мне кажется, что этот процесс уже идет тысячи лет. С момента появления первых технологий, ведь они делают нас умнее. Например, смартфон расширяет возможности мозга. Если тысячи лет назад смена парадигмы растягивалась на века, то теперь она меняется каждый год, если не быстрее. Это нельзя считать взрывом.

Я думаю, что к 2029 г. компьютеры достигнут уровня человеческого интеллекта, и этот интеллект сразу станет сверхчеловеческим. Talk to Books за полсекунды читает 600 млн предложений из 100 тысяч книг, а я несколько часов читаю одну! Смартфон по ключевым словам может осуществлять поиск в базе всех человеческих знаний. Поиск Google уже вышел за эти рамки и начал

частично использовать семантические возможности. Понимание семантики пока ниже человеческого уровня, но работает оно в миллиард раз быстрее, чем думает человек. При этом и программное обеспечение, и оборудование будут постоянно совершенствоваться.

М. Ф.: Вы известны также своими идеями по поводу продления жизни с помощью технологий. Расскажите немного об этом.

Р. К.: Во-первых, я верю в возможность слияния с интеллектуальной технологией, которую мы создаем. Согласно моему сценарию, в кровеносную систему можно будет вводить медицинских нанороботов, в частности, расширяющих возможности нашей иммунной системы. Я называю это третьим мостом к радикальному продлению жизни. Первый мост — это доступные в настоящий момент методы. Второй — совершенствование биотехнологий и возможность внесения изменений в жизненную программу. А нанороботы в числе прочего войдут к нам в мозг и позволят напрямую подключаться к виртуальной и дополненной реальностям. И что важнее всего, с их помощью мы соединим верхние слои нашего неокортекса с синтетическим неокортексом в облаке.

М. Ф.: В Google ведутся подобные разработки?

Р. К.: Проекты, над которыми работает наша команда, можно назвать примитивным моделированием коры головного мозга на базе тех сведений о неокортексе, которыми мы располагаем. И к началу 2030-х гг. у нас уже будет хорошая модель неокортекса.

Два миллиона лет назад черепная коробка людей была меньше, но росла для размещения увеличивающегося мозга. Что это нам дало? Даже будучи приматами, мы были довольно сообразительными, а теперь научились мыслить на более абстрактном уровне. Изобретение технологий, науки, языка и музыки отличает нас от

нынешних приматов. Но объем черепа не может увеличиваться до бесконечности, иначе исчезнет возможность рожать детей.

Это ограничение снимет расширение, которое появится в 2030-х гг. Мощность облачных технологий удваивается каждый год. А значит, будет развиваться небиологическая часть нашего мышления. Если посчитать, получится, что к 2045 г. человеческий интеллект возрастет в миллиард раз. Речь идет о настолько мощном преобразовании, что предвидеть его сейчас так же нереально, как заглянуть за горизонт событий.

Когда поисковик Google или программа Talk to Books достигнут человеческого уровня интеллекта, преимущество в скорости позволит им начать экспоненциальный рост. Это и есть технологическая сингулярность. Плавный старт, за которым следует экспоненциальный рост.

М. Ф.: Вы утверждаете, что сингулярность сильно повлияет на медицину и продолжительность человеческой жизни. В прошлом году, выступая в MIT, вы сказали, что в ближайшие 10 лет большинство людей получит то, что вы назвали скоростью убегания от старения (longevity escape velocity). Это действительно случится так скоро?

Р. К.: С точки зрения биотехнологий сейчас переломный момент. Люди считают, что лекарства будут разрабатываться в том же медленном темпе, что и раньше. По большому счету медицинские исследования велись наугад. Фармацевтические компании пробовали тысячи соединений в поисках того, что сможет оказать какое-то действие, вместо того чтобы понять, как работает управляющее нашей жизнью программное обеспечение, и перепрограммировать его.

Утверждение, что генетические процессы обеспечиваются программно, не метафора. Это строка данных, которая формировалась в эпоху, когда долгая жизнь индивидов противоречила ин-

тересам популяции, потому что были ограничены такие ресурсы, как еда. Эпоха дефицита кончилась, началась эпоха изобилия. Информация о различных аспектах биологии каждый год становится все доступнее. Например, первое секвенирование ДНК стоило миллиард долларов, а сейчас его цена близка к тысяче. И каждый год вдвое увеличивается наша способность не только собирать генетические данные, но и понимать, моделировать и, самое главное, программировать их.

Современные клинические применения можно сравнить с ручейком, но в следующие десять лет нас ждет наводнение. Уже появились сотни способов глубокого вмешательства в геном, которые ждут законодательного регулирования. Мы можем устранить результат сердечного приступа, физически омолодить сердце, используя перепрограммированные стволовые клетки. Мы умеем выращивать органы и успешно имплантируем их приматам. Иммунотерапия, по сути, перепрограммирует иммунную систему. По собственной инициативе иммунная система не начинает бороться с раком, потому что в процессе эволюции не была настроена на борьбу с заболеваниями, возникающими в более позднем возрасте. Но ее можно заставить рассматривать рак как патогенный фактор. Это открывает перед нами большие перспективы. Уже проводились испытания, в процессе которых практически все больные с четвертой стадией рака выходили в ремиссию.

За следующие десять лет медицина сильно изменится. При должном внимании к своему здоровью человек сможет достичь скорости убегания от старения и увеличить продолжительность жизни. Бессмертия это не гарантирует, потому что можно погибнуть, например, от несчастного случая, но песок не просто перестанет высыпаться из часов нашей жизни, а начнет прибывать. Еще через десять лет мы сможем обратить вспять процессы старения.

М. Ф.: Вас иногда критикуют за чрезмерный оптимизм. Видите ли вы в ИИ какие-то отрицательные аспекты? Несет ли он какие-то опасности?

Р. К.: О минусах я писал больше чем кто бы то ни было, причем за десятилетия до того, как свою обеспокоенность выразили Стивен Хокинг и Илон Маск. В книге «Эпоха чувствующих машин» есть большая дискуссия о минусах ГНР — генетики, нанотехнологий и робототехники (вещах, составляющих ИИ). По ее мотивам в январе 2000 г. Билл Джой опубликовал в журнале *Wired* свою знаменитую статью *Why the Future Doesn't Need Us* («Почему мы не нужны будущему»).

М. Ф.: Основой послужила цитата из Теда Казински, известного как Унабомбер?

Р. К.: На одну страницу я поместил цитату, которая звучит как обычное выражение озабоченности, а на следующей странице читатель обнаруживал, что это цитата из Манифеста Унабомбера. Я достаточно обстоятельно описал экзистенциальную опасность ГНР. В вышедшей в 2005 г. книге «Сингулярность уже близка» эта тема обсуждается еще более подробно.

Я оптимистично считаю, что как вид мы сохранимся. Технологии несут нам больше выгоды, чем вреда, но при этом не нужно глубоко копать, чтобы увидеть, какие разрушения принес мирный XX в. При том что в целом мир стал намного лучше. Например, за последние 200 лет на 95 % снизился уровень бедности, а общемировой уровень грамотности, который не дотягивал до 10 %, теперь превышает 90 %.

Но люди имеют тенденцию судить о ситуации по тому, насколько часто они слышат плохие новости. Был проведен опрос, в котором приняли участие 24 тысячи человек из 26 стран. Им предлагали оценить, как изменился уровень бедности в мире за последние 20 лет. Только 1 % опрошенных сказали, что этот показатель упал почти вдвое, то есть дали корректный ответ. По мнению 87 %, ситуация ухудшается. Такая предрасположенность к негативному взгляду на мир обусловлена эволюционно — десять тысяч лет назад легкий шелест листьев мог означать приближение хищника.

М. Ф.: Но ведь есть разница между реальными и экзистенциальными рисками.

Р. К.: Мы неплохо справляемся с экзистенциальными рисками, которые несут в себе информационные технологии. Вспомните, как 40 лет назад группа ученых предвидела и перспективу появления новой биотехнологии, и ее опасность. Была проведена первая Асиломарская конференция. Этические стандарты и стратегии регулярно обновляются. И это работает. Число людей, пострадавших от преднамеренного или случайного злоупотребления биотехнологией, близко к нулю.

Этим мы обязаны законодательству. Но опасности существуют. Возникают все более мощные технологии, например такие, как методы CRISPR, позволяющие редактировать геном, значит, нужно разрабатывать новые стандарты.

Около полутора лет назад на очередной Асиломарской конференции впервые был поднят вопрос о принципах работы с ИИ и был разработан набор этических стандартов. На мой взгляд, он нуждается в доработке, причем безотлагательно.

М. Ф.: Сейчас многих волнует проблема контроля, или выравнивания. Беспокоит ли вас появление у суперинтеллекта целей, не соответствующих целям человека?

Р. К.: Проблема в том, что люди далеко не всегда могут согласовать свои собственные цели. Мы все время создаем инструменты для расширения своих возможностей. Технологии нужны, чтобы преодолевать наложенные на нас ограничения. Этой цели будет служить и ИИ. Никакого противостояния, которое так любят показывать в антиутопических фильмах, не возникнет. Произойдет слияние. Точнее, уже происходит. Телефон пока не является частью нашего тела или мозга, но мы уже не выходим без него из дома.

Я пошел учиться в MIT, потому что уже в 1965 г. там были компьютеры. Для доступа к компьютеру приходилось ехать на велосипеде

через весь кампус и показывать удостоверение личности. Всего полвека спустя мы стали носить компьютеры в карманах.

Можно проследить, как рост демократизации вел к улучшению взаимопонимания. Два столетия назад в мире была только одна демократическая страна. Столетие назад — полдюжины. Сейчас демократических принципов придерживается 123 из 192 признанных стран, то есть 64 %. Можно сказать, что демократия стала стандартом. В истории человечества это самое мирное время и улучшение всех аспектов жизни. И все благодаря технологиям, которые становятся более интеллектуальными и глубже интегрируются в нашу жизнь.

В современных конфликтах сила каждой враждующей группы увеличена с помощью технологий. И конфликты всегда будут, хотя я надеюсь, что более совершенные способы коммуникации увеличат человеческую способность к эмпатии, заложенную биологически. Благодаря технологиям мы больше узнаем о посторонних людях, начав сопереживать им, как близким.

М. Ф.: Какое влияние, с вашей точки зрения, ИИ окажет на экономику и на рынок труда? Ждет ли нас новая промышленная революция?

Р. К.: Давайте вспомним, как прошла последняя промышленная революция. Двести лет назад бизнес-модель гильдии ткачей перестала существовать из-за появления прядильных и ткацких станков. Они предсказывали, что появятся и другие машины и большинство людей потеряет работу, трудоустроены будут только избранные. Частично этот прогноз сбылся, так как действительно часть рабочих мест исчезла. Но занятость не снизилась, а выросла, так как увеличилось благосостояние общества.

Если бы в 1900 г., когда 38 % населения работало на фермах, а 25 % — на фабриках, опытный футуролог дал бы прогноз, что через 115 лет на фермах будет работать 2 % населения, а на фа-

бриках — 9 %, люди бы сильно испугались. Ведь такой прогноз означал бы, что их потомки останутся без работы.

Ведь футуролог не смог бы рассказать, что это будут за новые рабочие места, потому что их еще не изобрели. При этом, если в 1900 г. существовало 24 млн рабочих мест, сейчас их 142 млн, то есть в процентах от населения занятость выросла с 31 до 44 %. Если взять в качестве основы для сравнения доллар с постоянной покупательной способностью, получится, что в среднем по статистике за работу в час сейчас платят в 11 раз больше, чем в 1900 г. Рабочий год при этом сократился с 3000 до 1800 часов. А предлагаемые варианты рабочих мест стали намного интереснее. Я думаю, что следующая промышленная революция пройдет примерно по такому же сценарию.

М. Ф.: Вы уверены? Согласно статистике автоматизация, которая станет возможной благодаря успехам машинного обучения, может уничтожить более половины существующих рабочих мест. Потому что выполняемые операции настолько рутинны и предсказуемы, что не требуют ИИ человеческого уровня. Но ожидать, что люди успешно пройдут переподготовку с водителя такси на инженера-робототехника, не приходится.

Р. К.: Ваш прогноз базируется на модели «люди против машин». Но люди уже сделались умнее и получили способность выполнять высокоуровневые работы. Пока только с помощью внешних устройств, но со временем все эти вещи, делающие нас умнее, будут подключаться напрямую к мозгу.

Кроме того, постоянно развивается сфера образования. Если в 1870 г. в США было 68 000 студентов, то сейчас их 15 млн. Возьмите всех, кто работает в этой сфере, то есть преподавателей и обслуживающий персонал, и получите примерно 20 % рабочей силы, причем постоянно появляются новые вещи, которые кто-то должен делать. Лет шесть назад экономики мобильных приложений не существовало.

Еще необходимо учитывать грядущее изобилие. На ежегодном заседании Международного валютного фонда его директор-распорядитель Кристин Лагард спросила меня: «Чем обусловлен экономический рост? Информационные технологии невозможно есть, надевать и заселять».

Я ответил, что они смогут удовлетворить все типы номинально физиологических потребностей. Появится вертикальное сельское хозяйство в контролируемых ИИ зданиях с выращенными гидропонным способом фруктами и овощами. Появится клонированная мышечная ткань, которая при низких затратах обеспечит людей высококачественной пищей без химикатов и убийства животных. Уровень дефляции в сфере информационных технологий составляет 50 %; цена на вычисления, коммуникации и генетическое секвенирование ежегодно уменьшается вдвое. И эта дефляция будет направлена на создание традиционных и необходимых продуктов.

М. Ф.: То есть, с вашей точки зрения, такие технологии, как 3D-печать, роботизированные фабрики и сельское хозяйство, помогут снизить расходы практически на все?

Р. К.: Именно так. В 2020-х гг. с помощью 3D-печати будет производиться одежда. Мы по разным причинам еще не совсем там, но по крайней мере движемся в правильном направлении. Другие необходимые нам физические объекты будут распечатаны на 3D-принтерах, включая модули, с помощью которых за считанные дни можно будет построить здание. В конечном счете, информационные технологии, контролируемые ИИ, упростят производство всего, что нам требуется.

Солнечной энергией будет легче пользоваться благодаря применению глубокого обучения для создания более качественных материалов, в результате чего стоимость как накопления, так и сбора энергии быстро снижается. Общее количество солнечной энергии удваивается каждые два года, и такая же тенденция существует

с ветровой энергией. Возобновляемая энергия в настоящее время составляет всего около пяти циклов, по два года на удвоение, от удовлетворения ста процентов наших потребностей в энергии, и к этому времени она будет использовать одну тысячную часть энергии от солнца или ветра.

Кристин Лагард возразила, что есть еще один ресурс, который невозможно получить с помощью информационных технологий, — это территория, и наша планета уже перенаселена. Я ответил, что ощущение перенаселения создается от всеобщей потребности жить и работать в городе, которая уходит в прошлое из-за доступности виртуального общения. При этом 95 % Земли пустует.

До наступления 2030-х гг. мы сможем обеспечить всему населению планеты качество жизни выше того, которое мы сегодня представляем. Я выступил в TED с прогнозом, что к этому времени у нас будет универсальный базовый доход. Не очень большой, но достаточный.

М. Ф.: Фактически, в вашем варианте будущего проблема недостатка рабочих мест снимается, так как люди смогут не работать благодаря универсальному базовому доходу.

Р. К.: Я исхожу из предположения, что работа — путь к счастью. В будущем люди начнут работать не для обеспечения базовых нужд, а для того, чтобы внести свой вклад в общее дело и получать удовольствие.

М. Ф.: Но если человек работает для удовольствия, зачем платить ему деньги?

Р. К.: Экономическая модель уже начала меняться. Люди все больше понимают ценность образования. Обучение в вузе — это не работа, но оно считается полезной деятельностью. Если человек, не работая, сможет удовлетворять свои физиологические нужды,

мы будем двигаться к вершине пирамиды Маслоу. Собственно, если сравнить нашу жизнь с тем, что было в 1900 г., мы занимаемся этим непрерывно.

М. Ф.: Беспокоит ли вас тот факт, что первыми сильный ИИ могут создать китайцы? Ведь в Китае нет ограничений, связанных с конфиденциальностью, а количество населения позволяет получить намного больше данных, и потенциально выше вероятность появления нового Тьюринга или фон Неймана.

Р. К.: Я не считаю, что это игра, в которой выиграет только тот, кто первым придет к финишу. Если в Китае случится прорыв в области использования солнечной энергии или глубокого обучения, это будет полезно всем. Китай публикует так же много, как и США, и вся информация становится общедоступной. Посмотрите на программную библиотеку для глубокого обучения TensorFlow от Google, которая стала общественным достоянием. Мы тоже предоставили доступ к технологиям, лежащим в основе сервисов Talk to Books и Smart Reply, чтобы люди могли ими пользоваться.

Лично мне очень нравится, что Китай делает упор на экономическое развитие и предпринимательство, но я бы посоветовал им двигаться в направлении свободного обмена информацией. Я думаю, что это способствует прогрессу. Пример Кремниевой долины вдохновляет весь мир. Кремниевая долина стала символом предпринимательства, служения эксперименту и извлечения уроков из неудачного опыта. Словом, я думаю, что все идет прекрасно, и не считаю происходящее международным соревнованием.

М. Ф.: Но Китай — авторитарное государство, и все эти технологии могут использоваться военными. Если Google и DeepMind четко заявили, что они против даже опосредованного применения их технологий в военных целях, китайские компании Tencent и Baidu не делали таких заявлений.

Р. К.: Не нужно смешивать проблему применения ИИ в военных целях и авторитаризм. Это разные проблемы. Меня беспокоит политический режим в Китае и хотелось бы призвать Китай повернуться в сторону большей свободы информации и демократического управления. Мне кажется, что это благотворно скажется на их экономике.

Политические, социальные и философские проблемы по-прежнему остаются очень актуальными. Меня беспокоит не то, что ИИ вдруг начнет действовать самостоятельно. Меня беспокоит будущее человеческой популяции, которая уже построила технологическую цивилизацию. Наше совершенствование с помощью технологий будет продолжаться, поэтому лучший способ сделать ИИ безопасным — следить за тем, как мы сами себя ведем.

> *Мне нравится представлять мир, в котором вообще не нужно заниматься рутиной. Корзина сама избавляется от мусора, потому что так настроена умная инфраструктура, а о чистоте одежды заботятся роботы.*

Даниэла Рус

ДИРЕКТОР ЛАБОРАТОРИИ COMPUTER SCIENCE И ИИ В MIT (CSAIL), ПРОФЕССОР ЭЛЕКТРОТЕХНИКИ И COMPUTER SCIENCE В ШКОЛЕ ИНЖЕНЕРИИ ИМЕНИ ЭНДРЮ И ЭРНЫ ВИТЕРБИ, ДИРЕКТОР ОБЪЕДИНЕННОГО ИССЛЕДОВАТЕЛЬСКОГО ЦЕНТРА TOYOTA-CSAIL, ЧЛЕН КОНСУЛЬТАТИВНОГО СОВЕТА КОМПАНИИ TOYOTA RESEARCH INSTITUTE

Даниэла Рус — член ACM, AAAI и IEEE, а также Национальной инженерной академии и Американской академии искусств и наук. Специализируется на развитии автономной техники и повсеместной интеграции в нашу жизнь машин, помогающих с решением когнитивных и физических задач. Обладательница стипендии Макартура и лауреат премии Engelberger Robotics Award от Ассоциации робототехники 2017 г. Докторскую степень в области computer science получила в Корнеллском университете. Сотрудничала с труппой современного танца Pilobolus в рамках двух проектов на стыке технологий и искусства, работала над спектаклями Serafim («Серафим») и Umbrella («Зонтик»).

Мартин Форд: Как возник ваш интерес к ИИ и робототехнике?

Даниэла Рус: Еще ребенком я прочитала все популярные в то время научно-фантастические книги. В Румынии не было такого количества видеопродукции, как в США, но я с большим удовольствием смотрела сериал «Затерянные в космосе».

М. Ф.: Вы не первая, чья карьера началась благодаря интересу к научной фантастике.

Д. Р.: В сериале мне очень нравились крутой парнишка Уилл и робот. Тогда я не представляла, что когда-нибудь моя работа тоже будет связана с роботами. В школе я имела отличные оценки по математике и естествознанию, и когда настало время поступать в колледж, я поняла, что чистая математика для меня слишком абстрактна, и стала изучать computer science. Я специализировалась в информатике и математике. Второй специальностью стала астрономия, которая связывала меня с моими фантазиями о других мирах.

К концу обучения в бакалавриате я посетила лекцию Джона Хопкрофта — ученого, который получил премию Тьюринга. Там я услышала, что классической computer science пришел конец. Он имел в виду, что для большинства алгоритмов из теории графов, которые были созданы основателями вычислительных методов, уже нашли решения. Пора двигаться дальше. Наступало время роботов.

Идея показалась мне захватывающей, и над докторской диссертацией я начала работать под руководством Хопкрофта. Мне хотелось внести свой вклад в робототехнику, которая на тот момент была не развита. Единственным доступным нам роботом был манипулятор PUMA (programmable universal manipulation arm), мало напоминавший мои детские фантазии. Я изучала его работу с теоретической точки зрения. И пыталась реализовать собственные алгоритмы, чтобы перейти от симуляций к реальным системам. К сожалению, в то время у меня был доступ только к таким динамическим моделям кисти, как Utah/MIT hand и Salisbury

hand. Ни одна из них не могла развить силу и крутящий момент, которых требовали мои алгоритмы.

М. Ф.: Получается, между тем, что умели физические машины, и тем, что предлагали алгоритмы, был большой разрыв.

Д. Р.: Именно так. В то время я поняла, что машина — это замкнутое соединение тела и мозга, и заинтересовалась вопросами их взаимодействия.

Сейчас в моей лаборатории много нетрадиционных роботов: модульные, мягкие, а также созданные из пищи или бумаги. Мы рассматриваем новые типы материалов, форм, архитектур и различные представления о корпусе и активно разрабатываем математические основы, которые становятся общедоступной базой знаний.

М. Ф.: Вы директор одного из наиболее важных исследовательских проектов в сфере робототехники и ИИ. Чем именно занимается ваша лаборатория?

Д. Р.: Мы изобретаем новые компьютерные технологии, совершенствуем мир и готовим специалистов.

CSAIL — организация необычная. Студенткой я воспринимала ее как гору Олимп для технологий. И даже не могла представить, что попаду туда. Мне нравится думать о CSAIL как о месте, где рождается будущее.

Наша организация появилась после слияния Лаборатории computer science с Лабораторией ИИ. Вторая появилась в 1956 г., когда Марвин Минский собрал в Нью-Гемпшире своих друзей, месяц гулял с ними по лесу, не отвлекаясь на смартфон, и придумал новую область исследований — ИИ. Эта область связана с разработкой и проектированием машин, которые всё делают как люди. Ученые из CSAIL разрабатывают способы это реализовать. Свою принадлежность к этому сообществу я до сих пор считаю исключительной привилегией.

История computer science началась в 1963 г., когда у профессора MIT Боба Фано возникла сумасшедшая идея, что два человека могут использовать компьютер одновременно. Так начался Проект MAC. Это аббревиатура выражения Machine-Aided Cognition (познание с помощью машины), но шутили, что на самом деле это первые буквы фамилий Минского и Корби (Фернандо «Корби» Корбато) — лидеров тех двух лабораторий.

Многое из того, что сейчас воспринимается как обыденность, появилось благодаря исследованиям, которые велись в CSAIL. Это пароли, шифрование RSA, компьютерные системы с разделением времени, которые вдохновили на создание операционной системы Unix, оптическая мышь, объектно-ориентированное программирование, речевые системы, мобильные роботы с компьютерным зрением, движение свободного программного обеспечения — этот список можно продолжить. Сейчас CSAIL лидирует в создании облаков и облачных вычислениях, а также в демократизации образования с помощью массового открытого онлайн-курса (massive open online courses, MOOC) и в идеях, связанных с безопасностью, конфиденциальностью и многими другими аспектами вычислительной техники.

М. Ф.: Это крупная организация?

Д. Р.: CSAIL — крупнейшая исследовательская лаборатория в MIT. В пяти школах и одиннадцати отделах, на которые она подразделяется, работает более тысячи сотрудников. Сегодня в лаборатории 115 преподавателей, и у каждого из них своя большая мечта. Кто-то хочет усовершенствовать работу компьютеров, а кто-то — улучшить жизнь человечества с их помощью. Например, Шафи Гольдвассер хочет дать людям гарантированную возможность конфиденциальных переговоров через интернет. Тим Бернерс-Ли хочет создать Билль о правах, Великую хартию вольностей в интернете. Есть те, кто хочет обеспечить больным персонализированные и максимально эффективные методы лечения, и те, кто хочет увеличить возможности машин. Лесли Кельблинг мечтает создать лейтенант-коммандера Дейту, а Рас

Тедрейк — роботов, способных летать. Я же хочу создать роботов, способных менять форму, потому что было бы здорово иметь всепроникающих роботов, помогающих решать когнитивные и физические задачи.

Я чувствую вдохновение, думая о том, что всего 20 лет назад вычисления были прерогативой немногочисленных экспертов из-за специфики оборудования. И о том, как все изменилось десять лет назад, когда появились смартфоны, облачные вычисления и социальные сети. Сегодня для работы с компьютером не нужно быть экспертом. Люди настолько привыкли к ним, что не замечают, насколько зависимыми стали. Вы только представьте исчезновение интернета хотя бы на один день. Больше нет социальных сетей, связи по электронной почте, GPS, медицинской диагностики, цифровой музыки, онлайн-магазинов. И мне становится любопытно, как мог бы выглядеть наш мир с роботами.

М. Ф.: Как лаборатория на базе университета сохраняет баланс между исследованиями и коммерческими разработками, которые в итоге превращаются в продукты?

Д. Р.: Мы не берем заказы от компаний, а занимаемся обучением, после которого студенты могут заняться наукой, найти работу в сфере высоких технологий или стать предпринимателями. Мы поддерживаем все варианты. Это наш способ технологического предпринимательства. CSAIL стала колыбелью для сотен компаний, но все они работают автономно.

Продукты мы тоже не создаем, но всегда рады, когда наши исследования приводят к их появлению. Основная миссия лаборатории связана с работой на будущее. Хотя актуальные идеи также принимаются в разработку.

М. Ф.: Какие инновации в области робототехники нас ожидают?

Д. Р.: Робототехника уже поменяла наш мир. Появились дистанционное обучение, медицинская помощь, робота на производстве,

мониторинг объектов по датчикам, 3D-печать. Мне нравится представлять мир, в котором вообще не нужно заниматься рутиной. Корзина сама избавляется от мусора, потому что так настроена умная инфраструктура, а о чистоте одежды заботятся роботы. Транспорт становится таким же доступным, как вода или электричество. Виртуальные помощники экономят рабочее время, делая человека здоровее, а результат труда — качественнее.

М. Ф.: А когда я смогу, стоя в центре Манхэттена, вызвать беспилотное такси?

Д. Р.: Начну с того, что в некоторой степени технологии автономного вождения доступны уже сейчас. Современные решения подходят для определенных ситуаций четвертого уровня автономии (согласно определению Сообщества автомобильных инженеров, это предпоследний уровень до полной автономии). На низких скоростях в средах с низкой сложностью и малым количеством взаимодействий уже функционируют автомобили-роботы, доставляющие людей и посылки. Для Манхэттена они не подходят, но могут работать в общинах пенсионеров или бизнес-кампусах.

Установленные в беспилотных автомобилях датчики не очень надежны в плохую погоду. Нужно научить системы ездить в пробках и продумать устройство смешанной среды человек/машина. Для достижения пятого уровня, на мой взгляд, может потребоваться еще десятилетие. Пока коммерческое применение автономных автомобилей возможно только локально.

М. Ф.: То есть мы говорим об услуге, которая будет ограничена проложенными маршрутами или хорошо размеченными районами?

Д. Р.: Не обязательно. Недавно вышла статья с результатами тестирования одной из первых систем, способных ездить по загородным дорогам. Кроме того, 10 лет — это большой срок. Вспомните, как 20 лет назад главный научный сотрудник Xerox PARC Марк Вейзер говорил о повсеместном использовании вычислительной техники, а его считали мечтателем.

В отношении технологий я предпочитаю быть оптимистом. На мой взгляд, они имеют огромный потенциал для объединения людей и расширения их прав и возможностей. Но сначала нужно усовершенствовать науку и технику, а также создать образовательные программы, которые научат людей пользоваться технологическими достижениями. Или подойти к вопросу с другой стороны — продолжить развивать технологическую сторону, чтобы машины сами начали адаптироваться к людям.

М. Ф.: Роботов, которые могут приносить пиво из холодильника, пока не существует.

Д. Р.: К сожалению, да. Мы пока достигли больших успехов в навигации, а не в манипулировании. Это два основных типа функциональных возможностей роботов. Прогресс в области навигации обеспечило появление датчика LIDAR, который представляет собой лазерный сканер. Появилась возможность использовать алгоритмы, которые не работали с сонаром, и это стало переломным моментом. Начали прогрессировать такие вещи, как картография, планирование и локализация, что вызвало большой энтузиазм в сфере автономного вождения.

А вот процесс манипулирования по большей части остался на том же уровне, что и 50 лет назад. Жесткие промышленные манипуляторы с двузубыми клешнями — это не то, что нам нужно. Понемногу мы начинаем задумываться о том, что такое робот. В частности, разрабатываются роботы с мягкими манипуляторами.

Традиционная рука робота с металлическими пальцами способна только к так называемому «жесткому контакту». Палец кладется на объект, который нужно взять, и в эту точку прикладывается сила и момент. При этом нужно знать точную геометрию этого объекта и точно рассчитать, в какую точку его поверхности положить пальцы, чтобы все силы и моменты уравновешивались и могли компенсировать внешние силы и моменты. В технической литературе это называется проблемой силового и геометрического замыкания.

Люди берут предметы по-другому. Попытайтесь поднять чашку ногтями. В случае мягких пальцев знать точную геометрию объекта не обязательно, потому что контакт возникает на большей площади.

Сделав корпус машины более функциональным, мы сможем контролировать ее с помощью алгоритмов различного типа. Я надеюсь, что мягкая робототехника продвинет вперед область, которая находилась в состоянии стагнации много лет. Но пока, несмотря на большой прогресс, нам еще далеко до возможностей, которыми обладают природные системы, то есть люди или животные.

М. Ф.: В каком направлении, на ваш взгляд, следует двигаться для создания сильного ИИ? И сколько времени могут занять эти разработки?

Д. Р.: Работа над ИИ ведется более 60 лет. Если бы основатели отрасли смогли увидеть то, что сейчас считается крупными достижениями, они были бы разочарованы. Большого прогресса мы пока не добились. И я не думаю, что в ближайшем будущем имеет смысл говорить о сильном ИИ.

Когда в массмедиа говорят об ИИ, зачастую авторы не понимают, что это такое, имея в виду машинное обучение или даже глубокое обучение. Эта тема очеловечена из-за терминов «интеллект» и «обучение», которые ассоциируются с людьми. Однако в процессе обучения, чтобы опознавать кофейные чашки, система как бы говорит себе: «Данная совокупность пикселов, которая на этой фотографии представляет кофейную чашку, такая же, как на других изображениях, помеченных как кофейные чашки». При этом она не имеет ни малейшего представления о том, что такое кофейная чашка. Разрыв между человеческим и машинным интеллектом огромен. Проблема понимания интеллекта сейчас одна из самых актуальных. Решение, скорее всего, лежит на стыке нейробиологии, когнитивистики и computer science.

М. Ф.: Может ли произойти какой-то прорыв, который позволит резко двинуться вперед?

Д. Р.: Это возможно. Например, мы сейчас пытаемся понять, можно ли создать робота, который будет адаптироваться к людям. Ищем способы распознавать и классифицировать мозговую деятельность.

Классификации подлежит реакция человека на сигнал, который называется «потенциалом, связанным с ошибкой». Этот сигнал мозга есть у всех людей. Его позволяют зафиксировать электроды в шлеме ЭЭГ. Можно представить ситуацию, когда оператор-человек наблюдает за работой роботов, и если он замечает сделанную роботом ошибку, то через специальное приложение передается сигнал, и робот корректирует свое поведение. Подобный проект уже начал свою работу.

Шлем ЭЭГ состоит из 48 электродов, установленных на голове человека. Эта механическая система напоминает о том, что когда-то компьютеры использовали механические переключатели. При этом у нас есть возможность с помощью инвазивных процедур подключаться к нейронам на уровне нервных клеток. То есть в мозг втыкаются зонды, позволяющие точно определить активность на нервном уровне. Сейчас существует большой разрыв между тем, что можно сделать извне и инвазивно. И мне интересно, произойдет ли в какой-то момент, согласно закону Мура, какой-то прорыв, который позволит воспринимать мозговую активность с более высоким разрешением.

М. Ф.: Какие опасности несет в себе эта технология? Например, не приведет ли она к массовой безработице? Или вы считаете, что люди сумеют адаптироваться?

Д. Р.: Ситуация на рынке труда менялась на протяжении всей истории. Современные технологии позволяют автоматизировать рутинную физическую работу с повторяющимися операциями и работу с данными. Меня это очень вдохновляет, потому что технологии могут освободить наше время для более интересных задач. Например, я обсуждала с физиотерапевтами автономную инвалидную коляску. Они очень рады ее появлению, потому что

до сих пор врачу приходилось идти к кровати пациента, усаживать его в инвалидную коляску и везти в тренажерный зал, где проходили занятия. Через час пациента нужно отвезти обратно. То есть изрядное время уходило на перемещения. А теперь представьте, что физиотерапевт может все время находиться в спортзале, куда пациентов доставляет автономное инвалидное кресло. Кроме того, гораздо проще проанализировать, какие профессии могут исчезнуть, чем представить, какие профессии могут появиться. Например, в XX в. в Соединенных Штатах число работающих по найму в сельском хозяйстве упало с 4 до 2 %. В начале века никто не догадывался, что так случится. При этом вспомните, что всего 10 лет назад, когда начался расцвет компьютерной индустрии, никто не прогнозировал уровень занятости в социальных сетях, в магазинах приложений, в облачных вычислениях, и даже в таких вещах, как консультирование в колледжах.

М. Ф.: То есть вы верите, что новые рабочие места компенсируют исчезновение старых?

Д. Р.: К сожалению, поводы для беспокойства тоже есть. Один из них — качество рабочих мест. Иногда внедрение технологии снижает требования к специалистам. Например, раньше таксистам требовались хорошая память и умение ориентироваться в пространстве. С появлением GPS нужда в этих навыках отпала. Профессия таксиста стала доступна большему количеству людей, что негативно отразилось на размере заработной платы. Во-вторых, непонятно, смогут ли люди пройти обучение, для того чтобы занять хорошо оплачиваемые рабочие места, появившиеся в результате развития технологий. Вы не представляете, сколько раз в день я слышу: «Нам нужны ваши студенты». Специалисты по машинному обучению и ИИ крайне востребованны. В итоге, с одной стороны, есть много рабочих мест, а с другой — множество людей ищет работу. Поэтому нужны программы переподготовки.

Я твердо верю в то, что переобучиться может любой. Мой любимый пример — запущенный пару лет назад в Кентукки проект

BitSource. Эта компания занимается перепрофилированием шахтеров, потерявших работу из-за закрытия неприбыльных шахт. Многие уже получили гораздо лучшую, более безопасную и более приятную работу. Это пример, который наглядно демонстрирует, что правильные программы и правильная поддержка могут реально помочь людям в переходный период.

М. Ф.: Вы говорите только о перепрофилировании или считаете, что нужно коренным образом менять всю систему образования?

Д. Р.: В XX в. грамотным считался человек, который умел читать, писать и знал арифметику. В XXI в. это определение нужно расширить, добавив навыки работы с компьютером. Если в школах начнут обучать созданию вещей с помощью программирования, возможности учеников вырастут.

Кроме того, нужно поменять отношение к обучению. Сегодня большинство сначала учится, а в какой-то момент начинает работать. Мне кажется, что с развитием технологий более правильным окажется параллельный подход. Человек всегда должен быть готовым к приобретению новых навыков и их применению в процессе обучения.

М. Ф.: В некоторых странах ИИ становится стратегическим направлением. Они принимают четкую промышленную политику, ориентированную на робототехнику и ИИ. В частности, активно инвестирует в эту область Китай. Не рискуем ли мы проиграть в этой гонке?

Д. Р.: Мне нравится то, что происходит в сфере ИИ по всему миру. Огромные инвестиции в эту сферу делают Китай, Канада, Франция, Великобритания и десятки других стран. Многие связывают с ИИ свое будущее, и я думаю, что США следует поступить так же. Нужно разглядеть потенциал, который несет ИИ, и увеличить поддержку и финансирование этой сферы.

> *Кто-то должен регулировать ИИ. Главное — не останавливать его использование, не закрывать ящик Пандоры и не откладывать применение новых технологий, пытаясь повернуть время вспять.*

Джеймс Маника

СТАРШИЙ ПАРТНЕР В MCKINSEY & COMPANY И ПРЕДСЕДАТЕЛЬ MGI, ЧЛЕН КОМАНДЫ AIINDEX.ORG, СОТРУДНИК DEEPMIND, НАУЧНЫЙ СОТРУДНИК INITIATIVE ON THE DIGITAL ECONOMY В MIT

Автор книг и статей Джеймс Маника принимает участие в организованной Стэнфордским университетом программе 100-летнего исследования ИИ. Член Совета Оксфордского института интернета, Совета по международным отношениям, Совета консультантов по цифровой экономике, Национального консультативного совета по инновациям, фонда Макартуров, фонда Хьюлетта и организации Markle Foundation. В 2012–2016 гг. занимал пост заместителя председателя Совета по глобальному развитию при Белом доме.

Мартин Форд: Как возник ваш интерес к робототехнике и ИИ?

Джеймс Маника: Наука всегда вдохновляла меня, отчасти потому, что мой отец был первым чернокожим из Родезии, получившим стипендию по программе Фулбрайта. Приехав в США в начале 1960-х гг., он посетил NASA и увидел с мыса Канаверал, как ракеты взлетают в небо. После возвращения отец много рассказывал о науке, космосе и технологиях. Я стал создавать модели самолетов и машин из всего, что попадалось под руку.

Когда я поступил в университет, Зимбабве уже была независимой. Я получил степень бакалавра в области электротехники с уклоном в математику и computer science. Именно тогда приглашенный исследователь из Университета Торонто привлек меня к проекту по нейронным сетям. Я узнал о методе обратного распространения Румельхарта и использовании сигмоид в качестве функций активации.

Я очень старался и в результате получил стипендию Родса для поступления в Оксфордский университет. Там я работал в Исследовательской группе по программированию под руководством Энтони Хоара — изобретателя алгоритма сортировки. Магистерскую диссертацию по математике я защищал на материале алгоритмов. От идеи стать космонавтом я отказался, но думал, что работа в сфере робототехники и ИИ приближает меня к науке о космосе.

Я попал в Исследовательскую группу по робототехнике в Оксфорде, где фактически велась работа над ИИ, но в то время этот термин воспринимался негативно. Был конец так называемой «зимы ИИ» после неоправданных ожиданий. Наша деятельность называлась как угодно — машинное восприятие, машинное обучение, робототехника или просто нейронные сети. Сейчас ситуация ровно противоположная. Все хотят добавить термин ИИ в описание своей работы.

М. Ф.: Когда все это происходило?

Д. М.: Работу над докторской диссертацией я начал в 1991 г. и вместе с Эндрю Блейком и Лайонелом Тарасенко работал над нейронными сетями. Майкл Брэйди, теперь сэр Майкл, работал над машинным зрением. Моим руководителем стал Хью Даррант-Уайт, работавший над распределенным ИИ и роботизированными системами. Вместе мы создали несколько автономных транспортных средств и написали об этом книгу.

Мне довелось сотрудничать с командой Лаборатории реактивных двигателей NASA, работавшей над марсоходом. Их интересовали системы машинного восприятия. Космические впечатления!

М. Ф.: Написанный вами код на самом деле используется на марсоходе?

Д. М.: Да. Я работал с группой Man Machine Systems в Лаборатории реактивного движения в Пасадине, штат Калифорния, как один из приглашенных ученых, разрабатывавших алгоритмы машинного восприятия и навигации. Некоторые из этих алгоритмов сейчас применяются в модульных и автономных системах транспортных средств и не только.

Именно в этот период зародился мой интерес к ИИ. Я обнаружил, насколько увлекательна такая вещь, как машинное восприятие. Мы разрабатывали алгоритмы машинного обучения для распределенных и многоагентных систем. Эти алгоритмы должны были понимать окружающую среду и автономно создавать ее модели, даже если никогда не видели этой среды раньше, обучаться по ходу дела.

Многие из моих разработок нашли применение в распределенных сетях и сборе и обобщении данных от различных датчиков. Мы строили системы машинного обучения, используя комбинацию байесовских сетей, придуманных Джудой Перлом, с фильтрами Калмана и другими алгоритмами оценки и прогнозирования. Системы должны были извлекать данные из окружающей среды и широкого спектра источников различного качества, обучаться

на них и делать прогнозы. В незнакомых средах они должны были уметь строить карты, собирать информацию о том, что их окружает, и дальше принимать решения, как интеллектуальные системы.

Сейчас я занят в сфере бизнеса, но все равно продолжаю следить за всем, что происходит в области машинного обучения и ИИ.

М. Ф.: Насколько я знаю, свою техническую карьеру вы начали с преподавательской деятельности?

Д. М.: Да, в колледже Баллиол в Оксфорде я преподавал математику и информатику, а также знакомил студентов с робототехникой.

М. Ф.: А как вы перешли к консалтингу по вопросам управления и бизнеса в компании McKinsey?

Д. М.: Это была случайность. Я получил предложение от McKinsey присоединиться к ним в Кремниевой долине и подумал, что это интересный поворот в моей карьере.

В то время, как и многие мои коллеги, например Бобби Рао, я интересовался системами, с которыми можно было участвовать в соревновании беспилотных автомобилей DARPA Grand Challenge. Многие наши алгоритмы допускали применение к автономным транспортным средствам, и эти соревнования были одним из немногих мест, где их можно было использовать. Все мои друзья тогда переезжали в Кремниевую долину. Бобби работал в Беркли вместе со Стюартом Расселом и другими, поэтому я решил принять предложение McKinsey и переехать в Сан-Франциско. Это позволило быть рядом с Кремниевой долиной и с местом, где происходили различные мероприятия.

М. Ф.: Какую роль вы сейчас играете в McKinsey?

Д. М.: Я работаю с новаторскими технологическими компаниями в Кремниевой долине и провожу исследования на стыке технологий, изучая их влияние на бизнес и экономику. Как председатель

MGI, я исследую не только технологии, но и макроэкономические и глобальные тенденции. У нас замечательные научные консультанты, среди которых есть и экономисты. Это Эрик Бринолфссон, Хэл Вариан и даже нобелевский лауреат Майк Спенс. Раньше с нами сотрудничал еще один нобелевский лауреат Роберт Солоу.

Мы следим за передовыми технологиями и прогрессом в сфере ИИ. Я поддерживаю контакты и сотрудничество с Эриком Хорвицем, Джеффом Дином, Демисом Хассабисом и Фей-Фей Ли, кроме того, учусь у легендарной Барбары Грош. В то время как я пытаюсь быть ближе к технологиям и науке, мои коллеги исследуют влияние этих технологий на экономику.

М. Ф.: Последние несколько лет глубокое обучение развивалось очень быстро. Как вы считаете, это путь к мечте или же не стоит многого ожидать от этого метода?

Д. М.: Мы только открываем для себя возможности глубокого обучения, нейронных сетей, обучения с подкреплением и переноса обучения. Запас того, что могут дать эти методы, огромен. Они уже помогают решать такие задачи, как классификация изображений и объектов, обработка естественного языка и генеративный ИИ, который прогнозирует и синтезирует последовательности для речи, изображений и прочего. Большой прогресс ожидается в так называемом узком ИИ, то есть в решении конкретных задач.

Для сравнения можно посмотреть на темпы разработки общего и сильного ИИ. Мы двигаемся быстрее, чем раньше, но необходим ряд прорывов. Впечатляет работа Джеффа Дина и других сотрудников Google Brain по автоматическому обучению машин. Но возможности машинного обучения ограничены.

М. Ф.: В чем их ограничения?

Д. М.: Маркированные данные не всегда доступны. Иногда их приходится создавать вручную, что занимает много времени и не гарантирует отсутствия ошибок. Для этого компании, занимаю-

щиеся беспилотными автомобилями, нанимают сотни людей. Появляются методы, позволяющие без этого обойтись. Например, предложенный Эриком Хорвицем метод in-stream supervision, присваивающий метки неявным образом непосредственно в процессе деятельности. Или генеративные состязательные сети (generative adversarial networks, GAN), которые представляют собой обучение с частичным привлечением учителя. Они генерируют полезные данные таким образом, чтобы уменьшить потребность в наборах данных, промаркированных людьми.

Вторая проблема состоит в количестве данных. Неудивительно, что больший прогресс достигнут в области машинного зрения, потому что в интернет ежедневно выкладывается множество изображений и видео. При этом доступность данных в некоторой степени уменьшают нормативные акты, конфиденциальность, безопасность и другие факторы. Этим частично объясняется разный темп прогресса в разных странах. Есть стандарты использования данных, которые упрощают доступ к сведениям из сферы здравоохранения. Именно поэтому Китай результативнее нас применяет ИИ в геномике.

Ограничены универсальные инструменты и не решены общие проблемы ИИ. Например, предлагаются новые формы теста Тьюринга.

М. Ф.: А как работают новые тесты?

Д. М.: Соучредитель компании Apple Стив Возняк предложил так называемый кофейный тест. Претенденту на звание сильного ИИ нужно войти в случайный дом, найти кофе-машину, приготовить кофе и разлить его по чашкам. Тест выглядит тривиально, но для его прохождения нужно принимать сложные универсальные решения множества проблем общего характера. Мне кажется, что это оптимальная форма теста Тьюринга.

Но я хотел бы вернуться к вопросу об ограничениях, так как осталась за кадром проблема, связанная с предвзятостью. ИИ-сообщество имеет полярные мнения по этому вопросу. Одна из точек зрения

заключается в том, что машины будут обладать меньшей степенью предвзятости, чем люди. Когда судья принимает решение об освобождении под залог, применение алгоритма может нивелировать человеческие предрассудки и даже избежать влияния суточных ритмов. Вспомните работу Марианны Бертран и Сендхила Малленатана, в которой рассматривались ответы на идентичные рабочие резюме, присланные представителями различных расовых групп.

М. Ф.: Многие надеются, что ИИ не унаследует человеческую предвзятость.

Д. М.: Предвзятость может появиться в данных как на стадии их сбора, так и за счет избыточной или недостаточной выборки. Эта проблема наглядно проявляется в кредитовании и уголовных делах, и велика вероятность, что в любой набор данных уже встроены масштабные предубеждения, причем непреднамеренно. Об этом много писали в своих работах Джулия Ангвин и ее коллеги из организации ProPublica и стипендиат фонда Макартуров Сендхил Малленатан. В итоге был сделан интересный вывод, что алгоритмы не способны удовлетворять различным определениям справедливости одновременно. Возникла важная проблема: определить, что такое справедливость.

Машины могут как помочь в преодолении предвзятости, так и усугубить ситуацию. К счастью, в этой области мы потихоньку продвигаемся вперед. Меня радует работа Сильвии Кьяппы из компании DeepMind, построенная на условной справедливости и причинно-следственных моделях.

М. Ф.: Но разве данные могут быть свободными от человеческих предубеждений, если они собираются в процессе обычной интернет-деятельности пользователей?

Д. М.: Я приведу пример, демонстрирующий, как возникает эта проблема. Известно, что отдельные районы патрулируются полицией более интенсивно, чем остальные. Это позволяет собрать о них больше данных.

Различия в выборке данных между строго охраняемыми и практически не патрулируемыми районами влияют на прогноз криминальной обстановки. Сбор данных может быть вполне объективным, но прогноз окажется далеким от реальности из-за чрезмерной выборки в одном районе и недостаточной выборки в другом.

Или рассмотрим такую процедуру, как выдача кредита. Есть люди, которые активно используют кредитные карты и совершают электронные платежи, и, соответственно, доступно множество данных об их транзакциях. Избыточная выборка в данном случае идет человеку на пользу, так как о нем можно сделать более точный прогноз. При этом о человеке, который предпочитает платить наличными, данных мало, алгоритм дает менее точный прогноз, что может негативно повлиять на окончательное решение о выдаче кредита.

Присутствует эта проблема и в системах распознавания лиц, что продемонстрировали в своих работах Тимнит Гебру, Джой Буоламвини и др.

М. Ф.: А что вы думаете о других опасностях, связанных с ИИ? Об угрозе, исходящей от суперинтеллекта?

Д. М.: Есть много поводов для беспокойства. Несколько лет назад группа, в которую входили, в частности, Илон Маск и Стюарт Рассел, встретилась в Пуэрто-Рико, чтобы обсудить прогресс в области ИИ, а также связанные с этой сферой проблемы. По итогам Стюарт Рассел опубликовал статью, где перечислялись области, которые недостаточно исследуются. Сейчас этот список считается исчерпывающим и включает вопросы безопасности.

Например, вопрос, как остановить вышедшую из-под контроля машину или даже алгоритм? Может понадобиться так называемая большая красная кнопка. Но исследователи из DeepMind в серии простых двумерных видеоигр gridworlds продемонстрировали, что многие алгоритмы теоретически могут научиться отключать свои выключатели.

При этом мы можем не знать, что именно повлияло на неверное решение или ошибочный прогноз, выданный ИИ. Существует интерпретатор сложных нелинейных моделей LIME, который пытается определить, на какие наборы данных опирается обученная модель при составлении прогноза. Другой многообещающий метод — применение обобщенных аддитивных моделей (generalized additive models, GAM). В них складываются модели отдельных признаков, по которым можно отследить, как влияет на прогноз добавление каждого признака.

Далее — проблема обнаружения. Для ИИ-систем нет мониторинга и разведки, и нельзя узнать, что где-то происходит использование ИИ-системы в террористических целях.

К счастью, постепенно появляются организации, небезразлично относящиеся к этим проблемам. Например, консорциумом Partnership on AI рассматриваются как вопросы предвзятости и безопасности, так и различные виды экзистенциальных угроз. Сэм Альтман, Джек Кларк и другие члены OpenAI хотят добиться выгоды от ИИ для всего человечества.

М. Ф.: Как вы относитесь к проблеме выравнивания, о которой предупреждают Илон Маск и Ник Бостром?

Д. М.: Я думаю, что специалисты, которые будут заниматься этим вопросом, нужны. Но сейчас волноваться об этом не имеет смысла, потому что если суперинтеллектуальные машины и появятся, то очень нескоро. Мне нравится, что над этой проблемой думает такой крупный философ, как Ник Бостром. А обычным людям беспокоиться не о чем.

М. Ф.: Мне близка ваша точка зрения. Об этих вопросах должны думать эксперты. Инвестировать в эти вещи государственные ресурсы сейчас неоправданно.

Д. М.: Разумеется, это не политический вопрос. Но с людьми, которые утверждают, что поводов для беспокойства нет, потому

что подобного просто не может произойти, я категорически не согласен.

Просто есть более актуальные проблемы. Это вопросы безопасности, использования и злоупотреблений, объяснимости, предвзятости, а также вопросы влияния на экономику и рынок труда. Именно с этими проблемами нам предстоит иметь дело в ближайшие несколько десятилетий.

М. Ф.: Должно ли правительство регулировать вопросы применения ИИ?

Д. М.: Думаю, на данный момент у нас нет ни инструментов регулирования, ни нормативно-правовой базы.

Кто-то должен регулировать ИИ. Главное — не останавливать его использование, не закрывать ящик Пандоры и не откладывать применение новых технологий, пытаясь повернуть время вспять. Это было бы неправильно, потому что технологии приносят огромную социальную и экономическую выгоду. Нужно больше говорить, например, о проблеме производительности труда, которую могут решить ИИ-системы.

М. Ф.: Давайте перейдем к экономическим аспектам. MGI выпустил несколько отчетов о влиянии ИИ на рынок труда. Я занимался этой темой, и с моей точки зрения, мы находимся на пороге глобальных изменений в этой сфере. При этом многие экономисты считают, что поводов для беспокойства нет.

Д. М.: На самом деле это не так. Думаю, мы на пороге новой промышленной революции. Технологии окажут влияние на бизнес, инновации и прогнозирование, а в некоторых случаях выйдут за рамки когнитивных способностей человека. Наши исследования в MGI показывают, что влияние будет несомненно положительным.

Экономика преобразуется за счет повышения производительности и появления новых товаров, услуг и бизнес-моделей. Развитие

ИИ откроет перед людьми возможность реализации собственных амбиций в совершенствовании мира через участие в проектах, позволяющих по-новому взглянуть на общественные проблемы и технологии. А исчезновение ряда рабочих мест будет скомпенсировано появлением новых.

М. Ф.: То есть вы считаете, что итоговое влияние будет позитивным, несмотря на трудности перехода?

Д. М.: Спрос на рабочую силу будет всегда. Кроме того, ряд факторов спроса никуда не исчезнет в ближней и среднесрочной перспективе. К таким факторам, например, относится рост благосостояния во всем мире, благодаря которому все больше людей начнет принадлежать к среднему потребительскому классу. Также произойдет трансформация рабочих мест, потому что технологии смогут дополнить рабочий процесс.

Проблема в относительных величинах этих трех процессов — исчезновения рабочих мест, возникновения новых и трансформации. Согласно нашим исследованиям, количество появившихся рабочих мест превысит число исчезнувших. Разумеется, этот вывод основан на ряде допущений, но все равно остается вопрос: с какими проблемами придется столкнуться? Определить их сложнее, чем оценить влияние ИИ на бизнес и экономику как положительное.

М. Ф.: Но посмотрите на показатели производительности за последнее время. Они не очень хорошие, и с точки зрения макроэкономических данных никакого роста мы не наблюдаем. Фактически, по сравнению с другими периодами мы имеем уровень производительности ниже среднего. Вы считаете, что это своего рода интервал запаздывания перед началом роста?

Д. М.: Согласно отчету, недавно опубликованному MGI, для медленного роста производительности есть множество причин. В частности, последние 10 лет были периодом самой низкой капиталоемкости.

Нельзя забывать и о таком важном факторе, как спрос. Существует закон, согласно которому выпуск продукции должен поглощаться спросом на нее, и отсутствие спроса вредит росту производства, что негативно сказывается на показателях производительности, независимо от уровня технологий.

М. Ф.: Это важный момент. Если развивающаяся технология мешает росту заработной платы, тем самым уменьшая платежеспособность среднестатистического потребителя, спрос снижается.

Д. М.: Вы смотрите в корень. Это критический параметр, особенно в странах с развитой экономикой, где 55–70 % спроса обеспечивается расходами потребителей. Нужны люди, зарабатывающие достаточно, чтобы иметь возможность покупать выпускаемую продукцию. Хотя мне кажется, что тут имеет значение еще и отставание в технологиях.

В 1999–2003 гг. мы с Бобом Солоу рассматривали выявленный им еще в 1980-х гг. парадокс производительности: «Мы видим компьютеры повсюду, но только не в официальных цифрах роста производительности». Парадокс был разрешен в конце 1990-х гг., когда спрос стал достаточно высоким, но, что еще важнее, современные технологии стали применяться в крупных секторах экономики — розничной и оптовой торговле. Архитектура «клиент — сервер» и ERP-системы трансформировали бизнес-процессы, что привело к росту производительности.

Сейчас мы переживаем подобное снова. Цифровые технологии в виде облачных вычислений, электронной коммерции или электронных платежей присутствуют повсюду, при этом в течение нескольких лет производительность практически не росла. Но если систематически измерять долю цифровой экономики, выяснится удивительная вещь: с точки зрения активов, процессов и того, как люди работают с технологиями, эта доля не так уж велика. Я даже не рассматриваю ИИ или следующую волну технологий с повсеместной оцифровкой.

Электронная коммерция в розничной торговле составляет всего около 10 %, в основном через сайт Amazon. При этом мы привыкли считать этот сектор высокоцифровым, но значительного прогресса в нем нет.

Так что, возможно, нам придется снова обходить парадокс Солоу. Пока степень оцифровки различных секторов не станет по-настоящему высокой, национальный уровень производительности не вырастет.

М. Ф.: Получается, что мы пока даже не столкнулись с влиянием ИИ и автоматизации на экономику?

Д. М.: Именно так. И отсюда следует еще один важный вывод: в будущем нам потребуется рост производительности, причем даже больший, чем мы можем себе представить. И именно ИИ, автоматизация и цифровые технологии будут его стимулировать.

Посмотрим на экономический рост в странах большой двадцатки (где чуть более 90 % мирового ВВП) за последние 50 лет. Такой срок выбран потому, что у нас есть данные за весь этот период. Так вот, между 1964 и 2014 гг. средний рост ВВП, зависящий от производительности и трудовых ресурсов, составил 3,5 %. Из них на увеличение трудовых ресурсов приходилось 1,7 %. Прогноз на следующие 50 лет с учетом старения населения и других демографических проблем показывает, что показатель прироста трудовых ресурсов упадет с 1,7 до 0,3 %.

В результате получается, что, если в течение следующих 50 лет производительность не вырастет, нас ждет экономический спад.

М. Ф.: Это перекликается с прогнозами экономиста Роберта Гордона, который говорит о том, что экономического роста не будет[1].

[1] Вышедшая в 2016 г. книга Роберта Гордона *Rise and Fall of American Growth* («Взлет и падение американского развития») содержит крайне пессимистичные прогнозы по поводу будущего экономического роста в Соединенных Штатах.

Д. М.: Все-таки Боб Гордон задается вопросом, появятся ли инновации, сравнимые с электрификацией, способные реально стимулировать экономический рост. Но сам в них не верит.

М. Ф.: Я надеюсь, ИИ сможет стать такой инновацией?

Д. М.: Мы тоже на это надеемся! Ведь это технология общего назначения.

М. Ф.: MGI публикует отчеты о том, что происходит на рынке труда и как меняется заработная плата. Как вы определяете, допускает ли автоматизацию каждое рабочее место и какой процент рабочих мест подвержен риску исчезновения?

Д. М.: Я предпочитаю отдельно рассматривать рабочие места, которые могут исчезнуть, преобразоваться и появиться.

На исследованиях и отчетах об исчезающих рабочих местах спекулируют все кому не лень. Мы в MGI предпочитаем отталкиваться не от профессий, а от задач. Рассмотрев более 2000 задач по информации из различных источников, в том числе из базы данных O * NET и Бюро статистики труда, а также 18 видов навыков для их выполнения, мы попытались оценить, в какой степени для их автоматизации применимы существующие технологии. И пришли к выводу, что в ближайшие 15 лет в США автоматизацию допускает примерно 50 % видов деятельности.

М. Ф.: Получается, что уже сейчас можно автоматизировать половину того, что делают люди?

Д. М.: Именно так. Другой вопрос, как отдельные задачи сопоставляются с профессиями? Если подняться на этот уровень, окажется, что доля профессий, в которых можно автоматизировать более 90 % задач, составляет 10 %. Примерно в 60 % профессий автоматизируется около трети составляющих их видов деятельности. Эти профессии попадают в категорию рабочих мест, которые подвергнутся преобразованиям.

М. Ф.: Я помню, что пресса положительно оценила ваш отчет. Но ведь автоматизация трети задач означает, что там, где работали трое, справятся двое.

Д. М.: Да, я как раз собирался об этом сказать. Работу можно преобразовать множеством способов. Необязательно ждать момента, когда все задачи будут автоматизированы, следует реорганизовать рабочий процесс. Сокращение может коснуться большего числа людей, чем казалось сначала.

Поэтому в нашем исследовании были рассмотрены и другие соображения. Вопрос технической осуществимости это только первый из пяти факторов, которые нужно учесть.

Второй фактор касается стоимости разработки и развертывания технологий. Очевидно, что далеко не все вещи, которые технически возможны, воплощаются в жизнь. Например, строить электромобили мы могли и 50 лет назад, но когда это действительно произошло? Когда затраты на покупку, обслуживание и зарядку таких автомобилей стали достаточно разумными, чтобы потребители захотели их купить, а компании захотели заняться их производством. Стоимость развертывания зависит и от того, о замене какой работы — физической или когнитивной — идет речь. Как правило, автоматизация когнитивной работы требует программного обеспечения и стандартной вычислительной платформы, поэтому предельные издержки могут быстро снизиться. А вот для автоматизации физического труда требуются машины, стоимость которых снижается не так быстро.

Третий фактор — динамика спроса на рынке труда. Здесь учитывается качество и количество рабочей силы, а также связанный с этим уровень заработной платы. Давайте рассмотрим работу бухгалтера и садовника.

Технически проще автоматизируется часть работы бухгалтера, связанная с анализом и сбором данных. Садовник же в основном

выполняет физическую работу в крайне неструктурированной среде, где могут появляться непредвиденные препятствия. Поэтому техническая сложность автоматизации во втором случае выше. Теперь посмотрим на стоимость развертывания системы. В первом случае достаточно программного обеспечения на стандартной платформе, то есть предельные издержки стремятся к нулю. Для замены садовника нужна роботизированная машина с множеством движущихся частей, и затраты на развертывание в данном случае будут выше. Если рассмотреть количество и качество труда, а также динамику заработной платы, опять же предпочтительнее окажется автоматизировать профессию бухгалтера. Ведь садовник в США получает примерно восемь долларов в час, в то время как бухгалтер — около тридцати.

Все эти выкладки показывают, что на самом деле автоматизировать некоторые низкооплачиваемые рабочие места сложнее как с технической, так и с экономической точки зрения.

М. Ф.: Получается, у выпускников университетов есть повод для беспокойства.

Д. М.: Я не знаю. Зачастую сравнивают высокооплачиваемую работу с низкооплачиваемой, а высококвалифицированную с низкоквалифицированной, но непонятно, насколько корректно такое сравнение. Деятельность, которая допускает автоматизацию, нельзя однозначно сопоставить со структурой заработной платы или требованиями к навыкам.

Еще на вероятность автоматизации влияет такой фактор, как итоговая выгода. Есть области, в которых людей имеет смысл заменять машинами, причем не для уменьшения выплат, а для получения результата, который нельзя получить человеческими силами. Это работа, связанная с восприятием и прогнозированием. Сюда же можно будет отнести беспилотные транспортные средства, после того как они достигнут определенной точки развития.

Пятый фактор можно назвать социальными нормами, что включает в себя как потенциальную регуляцию, так и общественное признание. Отличным примером в данном случае может послужить отношение к беспилотным автомобилям. Сегодня общеизвестно, что пассажирские самолеты управляются человеком-пилотом менее 7 % времени. Но пассажиры не волнуются по этому поводу, так как попросту не видят, что происходит внутри кабины. А вот автомобиль, который движется сам по себе, вызывает недоверие.

В настоящее время активно изучается степень социальной приемлемости и комфорта человека при общении с машинами. В MIT исследуется, как это выглядит в разных возрастных группах, социальных условиях и странах. Например, в Японии присутствие машины среди людей более приемлемо, чем в ряде других стран. Представьте, что вместо врача в палату въезжает экран с диагнозом. Уверен, большинство будет против.

М. Ф.: И что в итоге будет с рабочими местами в целом?

Д. М.: При подготовке нашего последнего отчета были рассмотрены все факторы, которые я упоминал выше, и по результатам создан ряд сценариев. В среднем получается, что к 2030 г. в мире может исчезнуть до 400 млн рабочих мест. Это примерно 15 % от общего количества рабочей силы. В развитых странах показатель будет выше, чем в развивающихся.

М. Ф.: А как будут обстоять дела с новыми рабочими местами?

Д. М.: В ближайшие 20 лет нас ждет гарантированный спрос на работу. В частности, из-за глобального роста благосостояния, благодаря которому больше людей сможет потреблять товары и услуги. Кроме того, население стареет, что создает спрос на определенные виды работы. Пока непонятно, насколько высоко будут оплачиваться услуги по уходу за пожилыми людьми, но спрос на них гарантированно будет расти.

В MGI мы рассматриваем и другие факторы, например вероятность будущей адаптации к климатическим изменениям и будущей модернизации существующих систем и инфраструктуры. Если развитые страны начнут инвестировать в инфраструктуру, это поможет увеличить число рабочих мест.

Кроме того, будут появляться совсем новые профессии. Один из научных консультантов MGI, Ричард Купер из Гарварда, предложил интересный вариант анализа данных из Бюро статистики труда. Обычно там отслеживается около 800 профессий, список которых завершает строка «Прочее». Туда попадают профессии, которые на момент сбора данных не были определены. Например, если посмотреть статистику за 1995 г., в этой категории вы увидите профессию веб-дизайнера. Что интересно, эта категория растет быстрее остальных, потому что постоянно появляются новые профессии.

М. Ф.: Да, я часто слышал этот аргумент. Например, 10 лет назад не существовало профессий, связанных с социальными сетями.

Д. М.: Именно так! Если посмотреть с разбивкой по десятилетиям, в США 8–9 % — это новые рабочие места. Появятся профессии, связанные с проектированием и обслуживанием машин и роботов.

М. Ф.: Но, согласно статистике, большинство людей — продавцы, водители грузовиков, медсестры, учителя, врачи или офисные работники. Все эти профессии существуют веками.

Д. М.: Да, сейчас это действительно так, но со временем некоторые из этих профессий все равно исчезнут. Мы изучили, где в течение 200 лет наблюдалось самое большое сокращение рабочих мест. И сравнили, что происходило в период индустриализации с тем, что мы имеем сейчас, когда появился ИИ. Оказалось, что количество исчезающих профессий укладывается в рамки нормы. И даже если предположить какие-то крайние варианты, мы все равно укладываемся в диапазон изменений опыта прошлых лет.

Но нас больше волнует процесс переквалификации, чем вопрос, будет ли достаточно работы.

М. Ф.: То есть существует вероятность массового несовпадения требуемых и имеющихся навыков?

Д. М.: При первом взгляде на процесс перехода может показаться, что в США рабочих мест достаточно. Но если посмотреть географическое распределение, в каких-то местах их количество ничтожно мало.

Нужно учитывать и фактор заработной платы. Большинство исчезающих профессий — со средней заработной платой. При этом рабочие места будущего могут не очень хорошо оплачиваться. Поэтому следует или поменять рыночные механизмы оплаты труда, или иначе сформировать структуру оплаты.

При этом если машина будет помогать высококвалифицированному работнику, его заработная плата вырастет вместе с производительностью. Но если после автоматизации все операции, приносящие добавленную стоимость, будут выполняться машиной, то человеку останется неквалифицированная часть работы и низкая заработная плата.

Разумеется, идея дополнения человека машиной может привести к широкому диапазону результатов. При этом есть тенденция заострять внимание на положительной границе диапазона, игнорируя возможность максимально невыгодного для работников варианта. Кроме того, люди сталкиваются с необходимостью постоянного повышения квалификации, потому что дополняющие их машины все время развиваются.

М. Ф.: Я помню, как появление GPS повлияло на таксистов.

Д. М.: Да, изначально таксистам требовалось отличное знание всех городских закоулков и умение сокращать дорогу. Когда появились системы GPS, для этой работы оказалось достаточно навыка во-

дить машину. И на нее стало претендовать гораздо больше людей. Или операторы колл-центров, которым раньше требовались технические знания, а после внедрения сценариев оператором может работать любой, кто умеет читать. Если запрос клиента невозможно решить с помощью сценария, оператор переключает его на технического специалиста.

Существует множество вариантов работы, где в найме остается только неквалифицированный персонал, которому можно платить меньше.

М. Ф.: То есть вас в целом больше заботит влияние на заработную плату, а не проблема безработицы?

Д. М.: Проблема безработицы меня тоже беспокоит, потому что нельзя исключать вероятность плохого сценария. Но более актуальными мне кажутся проблемы перепрофилирования и переобучения и то, как мы будем поддерживать людей в переходный период.

М. Ф.: Выше вы отмечали, что потребительский спрос влияет на рост производительности.

Д. М.: Да. Это порочный круг, который еще больше снизит спрос на работу. Поэтому действовать нужно быстро. К сожалению, согласно нашим исследованиям, в большинстве стран с развитой экономикой снижаются затраты на обучение без отрыва от производства. И в ближайшем будущем это станет реальной проблемой.

Так называемая «активная поддержка рынка труда» не имеет отношения к обучению без отрыва от производства. Это поддержка, предоставляемая сокращенным работникам, которым приходится искать другую работу. Фактически это просчет, который мы допустили в последнем раунде глобализации.

Можно много говорить о том, насколько глобализация важна для производительности, экономического роста, потребительского

ассортимента и для продуктов. Но для рабочих она становится проблемой. При этом поддержка уволенных по сокращению штатов, по сути, отсутствует.

М. Ф.: Можно ли решить эту проблему введением универсального базового дохода?

Д. М.: Лично мне эта идея не нравится. Хорошо, что она обсуждается.

Работа придает смысл жизни, дает человеку чувство собственного достоинства, несет общественные и социальные последствия. Мне нравится цитата из отчета *Technology and the American economy*, изданного учрежденной президентом Линдоном Джонсоном Национальной комиссией: «Технологии приводят к исчезновению рабочих мест, но работа, которую нужно делать, никуда не исчезает».

М. Ф.: Вопрос в том, насколько высоко эта работа будет цениться на рынке труда.

Д. М.: Далеко не вся работа связана с рынком труда. Например, неоплачиваемая работа по уходу за детьми, больными и пожилыми, которой часто занимаются женщины. О ней не упоминают при обсуждении заработной платы и доходов. Необходимость в ней никогда не исчезнет. Вопрос в том, будет ли она признана работой и будет ли за нее выплачиваться какая-то компенсация.

Я предпочитаю условный доход, при котором получение выплат будет связано с каким-либо видом деятельности, подразумевающим инициативу, цель, достоинство и другие важные факторы. Потому что именно они определяют нашу сущность.

> *Я не уверен, что простым увеличением количества воспринимаемых данных можно достичь точности, необходимой для вождения на Манхэттене. Можно получить точность 99,99 %, но даже этого будет недостаточно.*

Гари Маркус

ОСНОВАТЕЛЬ И ГЕНЕРАЛЬНЫЙ ДИРЕКТОР СТАРТАПА GEOMETRIC INTELLIGENCE, ПРОФЕССОР ПСИХОЛОГИИ И НЕЙРОБИОЛОГИИ В НЬЮ-ЙОРКСКОМ УНИВЕРСИТЕТЕ

Гари Маркус — автор и редактор ряда книг, в том числе бестселлера Guitar Zero. Известен критикой глубокого обучения и утверждает, что для дальнейшего прогресса в ИИ-системы потребуется включить врожденные когнитивные структуры человека.

Мартин Форд: Вы написали книгу о несовершенстве человеческого мозга. Получается, вы не верите в то, что, полностью скопировав мозг, мы сможем подойти к сильному ИИ?

Гари Маркус: Полное копирование означает перенос как достоинств, так и недостатков. В то время как брать целесообразно только те вещи, которые люди делают лучше современных машин. На самом деле не имеет значения, насколько сильный ИИ будет похож на человека. Просто сейчас люди — единственные, кто умеет делать выводы и строить планы на основе множества данных и продуктивно их обсуждать, поэтому важно понять, каким образом они это делают.

Моя первая книга, опубликованная в 2001 г., называлась *The Algebraic Mind* («Алгебраический разум»). В ней нейронная сеть сравнивалась с мозгом. Я анализировал, что нужно сделать для ее усовершенствования, и приведенные там аргументы до сих пор актуальны. В следующей книге *The Birth of the Mind* («Рождение разума») я описал, как на генетическом уровне закладываются врожденные способности. Развивая идеи Ноама Хомского и Стивена Пинкера, я попытался понять, как врожденность выглядит с точки зрения молекулярной биологии и нейробиологии. Мне кажется, что эта информация тоже по-прежнему актуальна.

В 2008 г. вышла моя книга *Kluge: The Haphazard Evolution of the Human Mind* («Клуг: бессистемная эволюция человеческого разума»). Термин kluge означает временное решение, которое может быть неэффективным, не допускающим расширения и сложным в обслуживании. Именно так во многих отношениях выглядит человеческий разум. На материале множества дискуссий о том, оптимально ли устроены люди, я попытался с эволюционной точки зрения понять, почему мы не оптимальны.

М. Ф.: Наверное, потому, что эволюции приходится работать с уже существующей структурой? Вернуться в отправную точку и внести изменения в конструкцию невозможно.

Г. М.: Именно так. Большая часть книги посвящена сравнению структуры памяти с другими системами. Например, наша слуховая система близка к теоретическому оптимуму. То же самое можно сказать и о зрении — поразительно, но при благоприятных условиях человек может разглядеть даже фотон. При этом наша память далека от оптимума.

В компьютер можно быстро загрузить все произведения Шекспира, и он их никогда не забудет. Человеческая память не может похвастаться подобной емкостью и стабильностью. Наших предков интересовала в основном статистическая информация вида: «На горе больше еды, чем под ней». При таком подходе не имеет значения, в какие конкретно дни собирались статистические данные, важно помнить общую тенденцию.

В отличие от позвоночных, компьютеры используют адресную память, где каждой ячейке соответствует свой адрес, что позволяет хранить информацию практически бесконечно. Чтобы человек смог помнить так же, потребуется вносить изменения в слишком большое количество генов.

При этом компьютерные системы позволяют создавать гибриды, такие как поисковая система Google, которая сначала задействует контекстно-адресную память для поиска по ключевым словам, а затем с помощью адресной памяти определяет, где лежит нужная информация, и корректно отвечает на запросы.

М. Ф.: Можете рассказать более подробно?

Г. М.: Человеческую память могут активизировать различные факторы. Вплоть до позы. То, что человек запоминал стоя, ему будет проще вспомнить стоя. Или печально известный пример, когда человек готовится к экзамену в нетрезвом виде — вероятность успешной сдачи повышается, если он придет в аналогичное состояние.

Не похоже, чтобы наш мозг обладал механизмами адресации отдельных воспоминаний. Вместо этого человек посылает мозгу

запрос вида: «Напомни, как испечь пирог». И в ответ получает набор относящихся к данному вопросу воспоминаний, хотя понятия не имеет, где физически они хранятся.

При этом воспоминания человека имеют свойство размываться и перемешиваться. Поэтому очевидцы одного события могут давать разные показания. К сожалению, воспоминание о том, что произошло с вами в определенный момент, невозможно сохранить отдельно от более поздних размышлений или информации, которую вы увидели по телевизору или прочитали в газете.

М. Ф.: Это интересно.

Г. М.: В книге «Клуг» я утверждаю, что существует два вида памяти, и людям достался менее эффективный. Но нельзя перестроить ее с нуля, поэтому новые конструкции приходится строить поверх существующих. Это напоминает аргументы, которые Стивен Джей Гулд приводил в своей книге о большом пальце панды.

Наша память несет с собой и такие вещи, как предвзятость подтверждения. Люди склонны лучше помнить факты, которые согласуются с их точкой зрения. Компьютер же ищет все, что соответствует заданному критерию. При этом он умеет пользоваться оператором НЕ. Например, я могу найти в базе данных все понятия, не попадающие в заданный критерий. А человеческий мозг легко умеет искать только совпадения.

Или другой пример когнитивных искажений — эффект фокусировки. Скажем, ответы на два вопроса — насколько вы счастливы в браке и насколько вы довольны своей жизнью, будут зависеть от того, в каком порядке они будут заданы.

М. Ф.: Это напоминает эффект привязки, который иллюстрируется экспериментом Даниэла Канемана. Когда сообщенное людям случайное число влияет на их дальнейшие предположения.

Г. М.: Да, это вариация того, о чем я говорю. Если я попрошу вас посмотреть последние три цифры на банкноте, а затем спрошу, когда была подписана Великая хартия вольностей, с большой вероятностью увиденное повлияет на ваш ответ.

М. Ф.: Изначально вы работали над вопросами понимания человеческого языка, а затем основали стартап и помогали запуску ИИ-лабораторий в компании Uber.

Г. М.: Я ощущаю себя как Джозеф Конрад, который писал по-английски, хотя его родным языком был польский. Он понимал, как этот язык функционирует. Я пришел в ИИ из когнитивистики.

В детстве я занимался программированием, но к моменту поступления в аспирантуру стал больше интересоваться тем, как устроено мышление. Моим научным руководителем стал Стивен Пинкер — популяризатор науки, специалист в области экспериментальной психологии, психолингвистики и когнитивных наук. Мы исследовали, как дети знакомятся с прошедшим временем, а затем учились применять двухслойные и многослойные персептроны.

В 1986 г. Дэвид Румельхарт и Джеймс Макклелланд опубликовали статью *Parallel Distributed Processing: explorations in the microstructure of cognition* («Параллельно распределенная обработка: исследование микроструктуры познания»), в которой показывалось, что нейронную сеть можно научить, как ребенка, использовать прошедшее время английского языка. Проблема была в том, что дети делают другие ошибки. Мы предположили, что у детей работает гибридная система из формально применяемых правил и принципа, по которому работают нейронные сети.

М. Ф.: Вы говорите о том, что дети образуют прошедшее время неправильных глаголов, пользуясь правилом для обычных глаголов?

Г. М.: Да, дети иногда придают неправильным глаголам свойства правильных. Машинный анализ 11 000 фрагментов детской речи привел нас к гипотезе, что дети склонны использовать для образования прошедшего времени некое правило. С одной стороны, они добавляют стандартное окончание «-ed», но одновременно используют ассоциативную память: если в прошедшем времени человек употребляет глагол sing как sang, ему будет проще вспомнить, что глагол ring в прошедшем времени звучит rang. Но если встречается новое слово, которое не похоже на слова, слышанные ранее, для образования прошедшего времени будет использовано стандартное правило.

Дело в том, что нейронные сети очень хорошо распознают сходство, но испытывают трудности при распознавании вещей, которые ни на что не похожи, но подпадают под некое правило. Так было в 1992 г., и, по сути, ситуация до сих пор не изменилась. Работа большинства нейронных сетей определяется данными, при этом они не обладают высоким уровнем абстракции. Например, система распознавания объектов на изображениях приняла дорожный знак, покрытый наклейками, за холодильник с едой и напитками.

М. Ф.: Вы исследуете механизмы, отвечающие за понимание речи и обучение языку. Расскажите об экспериментах, которые вы проводили.

Г. М.: С 1999 г. изучая взрослых, детей и младенцев, я обнаружил, что люди хорошо подмечают общие тенденции. Например, семимесячные дети после двух минут прослушивания примеров искусственной грамматики обучались распознавать правила построения предложений. Прослушав предложения вида «la ta ta» и «ga na na», построенные по схеме ABB, младенцы замечали, что «wo fe wo» построено уже по другой схеме (ABA), в то время как предложение «wo fe fe» использует ту же конструкцию.

Критерием служила длительность взгляда. Оказалось, что после изменения схемы младенцы смотрели на экспериментатора дольше.

Получается, дети с самого раннего возраста умеют распознавать довольно глубокие языковые абстракции. Позднее другой исследователь показал, что этим свойством обладают и новорожденные.

М. Ф.: Насколько я знаю, вы снова заинтересовались ИИ благодаря суперкомпьютеру Watson фирмы IBM.

Г. М.: Я скептически относился к этому проекту и сильно удивился, когда в 2011 г. суперкомпьютер победил в телевикторине «Своя игра». Я снова заинтересовался сферой ИИ и, в конце концов, понял, чем был обеспечен успех Watson. Дело в том, что перед ИИ была поставлена более узкая задача, чем казалось на первый взгляд. Так происходит почти во всех случаях, когда ИИ демонстрирует впечатляющие успехи. В телевикторине примерно 95 % ответов представляли собой заголовки страниц «Википедии».

Примерно в то же время я начал писать для еженедельника *The New Yorker* статьи о нейробиологии, лингвистике, психологии, а также ИИ. Тогда, пять лет назад, я сомневался, что глубокое обучение сможет освоить повышение уровня абстракции и выстраивание причинно-следственных связей. Сейчас эти вопросы все еще не решены.

М. Ф.: В 2014 г. вы основали стартап Geometric Intelligence, который приобрела компания Uber, а вы после этого стали главой их ИИ-лаборатории. Расскажите, как все происходило.

Г. М.: В январе того года я решил основать собственную компанию. Пригласил замечательных людей, в том числе одного из лучших специалистов по машинному обучению — моего друга Зубина Гахрамани. За следующие несколько лет я многое узнал о машинном обучении. Мы искали более совершенные способы обобщения и обучали алгоритмы извлекать информацию из данных. Мы научили машины решать произвольные задачи, такие как распознавание символов из базы MNIST, используя вдвое меньший объем данных, чем методы глубокого обучения.

О наших успехах начали говорить, и в декабре 2016 г. стартап купила компания Uber. Я некоторое время работал в ней, помогая с запуском Uber AI labs. Сейчас я оттуда ушел, изучаю возможности использования ИИ в медицине и много думаю о робототехнике.

В январе 2018 г. я написал две статьи[1], а также пару вещей для платформы социальной журналистики Medium. В одной я рассказывал, что глубокое обучение не приведет нас к общему ИИ. Вторая статья была посвящена тому, что по крайней мере в биологии все системы начинаются с некой внутренней структуры. И для понимания мира важна врожденная структура, присутствующая в нашем мозге.

Данное от природы и полученное от воспитания часто противопоставляются. Но на самом деле эти вещи работают совместно. Природа дает нам механизмы обучения, которые позволяют использовать приобретаемый опыт интересными способами.

М. Ф.: Наличие структур показали эксперименты с младенцами, которые умеют распознавать лица.

Г. М.: Именно так. Это подтверждают мои эксперименты с восьмимесячными детьми, а недавно в журнале *Science* была опубликована статья, в которой утверждалось, что способность к логическим выводам появляется только после первого года жизни. Дело в том, что термин «врожденный» не означает появляющийся в момент рождения. Борода начинает расти после полового созревания. Большая часть человеческого мозга развивается уже вне матки, при этом развитие начинается относительно рано.

Жеребята почти сразу начинают ходить и обладают сложным зрением, позволяющим видеть препятствия. У человека подобные механизмы включаются в первый год жизни. Когда ребенок учится ходить, мы наблюдаем процесс созревания. Голова с полностью

[1] https://arxiv.org/abs/1801.00631.

развитым мозгом оказалась бы слишком большой и не смогла бы пройти через родовые пути.

М. Ф.: Получается, что, если сразу после рождения мы обладали бы способностью ходить, все равно пришлось бы ждать развития мышц.

Г. М.: Да, мы рождаемся не до конца развитыми, и многое из происходящего с нами в первые месяцы заложено генетически. Об обучении в этом случае речи не идет. Козленок уже через пару дней после рождения может спускаться по склону горы. Он не учится этому методом проб и ошибок. Я думаю, что геном человека передает шаблон того, как должен работать мозг, а ребенок потихоньку все это развивает. При этом в шаблон уже заложены механизмы обучения.

Почему-то есть тенденция создавать ИИ с минимумом предварительных знаний. Мне это кажется неразумным. Информацию о мире нужно сразу встраивать в системы ИИ, нежели разрабатывать их с нуля.

М. Ф.: Все врожденные вещи в мозге, вероятно, появились в результате эволюции. В ИИ-системах их можно или закодировать, или использовать эволюционный алгоритм для их автоматической генерации.

Г. М.: К сожалению, эволюция — это довольно медленный и неэффективный процесс. Для получения хороших результатов нужны триллионы организмов и миллиарды лет. Вряд ли в лабораторных условиях таким путем можно далеко продвинуться в разумные сроки.

Кроме того, первые 900 млн лет эволюции ничего особо захватывающего не происходило. Разные версии бактерий сменяли друг друга, что не представляет особого интереса. Затем процесс ускорился, появились позвоночные, млекопитающие, приматы

и, наконец, мы. Причиной ускорения стало, если можно так выразиться, множество накопленных подпрограмм и библиотечного кода. А чем больше подпрограмм, тем быстрее можно создавать более сложные вещи. Одно дело внести в мозг приматов ряд генетических изменений, чтобы получить человека, но с уровня бактерии подобный скачок невозможен.

Люди, которые используют для тренировки нейросетей эволюционные алгоритмы, часто начинают практически с нуля. Они пытаются развить отдельные нейроны и связи между ними, но лично я считаю, что в процессе биологической эволюции люди обладали очень сложными наборами генетических процедур. По сути, работать следует с существующими наборами генов, но в контексте эволюционного программирования реализовывать это пока никто не умеет.

Думаю, рано или поздно к этому все придет, но пока ученые играют в бога, создающего мир за семь дней.

М. Ф.: А как может выглядеть процесс встраивания врожденных качеств в ИИ-систему?

Г. М.: Этот вопрос состоит из двух частей: функциональная часть — что нужно сделать и механическая — каким образом.

В начале 2018 г. на базе собственных исследований и работ Элизабет Спелке из Гарварда я написал статью, где перечислил необходимые вещи[1]. В этот список попали манипулирование символами, представление абстрактных переменных и операции с ними, умение отличать значение слова (оно обозначает конкретный объект или тип объектов), понимание причинно-следственной связи, инвариантность перевода, представление о связанных траекториях движения объектов в пространстве и времени, осознание многообразия мира.

[1] https://arxiv.org/abs/1801.05667.

Глубокое обучение с подкреплением дает впечатляющие, но крайне нестабильные результаты. Например, в компании DeepMind ИИ обучили игре Breakout от Atari. И на первый взгляд ИИ отлично справился: нашел способ рикошетить от стен так, чтобы сбить максимальное количество кирпичей. Но стоит переместить ракетку вверх на три пиксела, и чудо закончится, ведь система не знает, что такое стена и рикошет.

М. Ф.: Для понимания объектов и концепций нужен более высокий уровень абстракции.

Г. М.: Именно так. Возможно, даже имеет смысл сразу встраивать определенные понятия, например «объект». Вспомните, как люди учатся распознавать цвета. В сетчатке глаза присутствует три типа цветовых рецепторов, чувствительных к разным участкам спектра. Дальше ребенок постепенно узнает, как называются цвета, которые он видит. Но без врожденной способности различать цвета вторая часть становится невозможной. Поэтому ИИ-системам может понадобиться врожденное представление о существовании объектов и о том, что их появление и исчезновение происходят неслучайным образом.

Представьте мир, в котором существует машина для телепортации, как в сериале «Звездный путь», и в любом месте в любой момент может появиться что угодно. На таких данных учиться невозможно. Изучать свойства объектов позволяет тот факт, что в пространстве и времени объекты перемещаются по связанным траекториям, и так сложилось за миллиарды лет эволюции.

М. Ф.: Можно ли создать сильный ИИ с помощью существующих в настоящее время инструментов и какие препятствия стоят на пути к нему?

Г. М.: Глубокое обучение я считаю инструментом, отлично подходящим для классификации шаблонов. Пока нет ничего эффективнее. Но оно не учит рассуждать и делать выводы на абстрактном

уровне. Не очень подходит для работы с языком в случаях, когда требуется реальное понимание. Плохо справляется с ситуациями, которые раньше не возникали, и в случаях с неполной информацией. Поэтому его необходимо дополнять другими инструментами.

В более широком смысле, накопленные людьми знания о мире можно записать символически с помощью математики или предложений естественного языка. И мы хотим объединить эту символическую информацию с другой информацией, получаемой путем субъективного восприятия.

Психологи говорят о связи между информацией, которая обрабатывается снизу вверх (взгляд на изображение), и обрабатываемой сверху вниз (интерпретация увиденного на базе личного опыта).

Современные системы глубокого обучения имеют дело с восходящей информацией. Они могут интерпретировать пикселы, но не объект.

В статье *Adversarial Patch* («Состязательный патч»)[1] демонстрируется способ обмануть систему глубокого обучения, добавив на изображение стикер. Система без проблем распознает изображенный на фото банан, а затем рядом наклеивается картинка, напоминающая выкрашенный в ядовитые цвета тостер. Человек сразу скажет, что это банан и забавная наклейка, в то время как система глубокого обучения уверенно опознает всю картинку как тостер.

Дело в том, что она пытается указать на наиболее заметный фрагмент изображения, поэтому ее внимание привлекает высококонтрастный, яркий тостер, а неяркий, однотонный банан она игнорирует.

Этот пример доказывает, что системы глубокого обучения получают только восходящую информацию, за обработку которой у человека отвечает затылочная доля. Они не умеют воспроизво-

[1] https://arxiv.org/pdf/1712.09665.pdf.

дить процессы, происходящие в лобных долях, то есть рассуждать о том, что на самом деле происходит.

Для создания сильного ИИ должны воспроизводиться оба процесса. На мой взгляд, следует добавить к глубокому обучению манипуляции с символами.

М. Ф.: Какая компания или проект ближе всего подошли к созданию сильного ИИ?

Г. М.: Меня впечатлил проект Mosaic, над которым работают в AI2. Это вторая попытка решить проблему, над которой бился Дуг Ленат, — как преобразовать человеческие знания в форму, понятную программному обеспечению. Множество информации нигде не зафиксировано. Например, никто специально не фиксирует факт, что тостеры меньше автомобилей.

М. Ф.: То есть требуется добавить формальную логику?

Г. М.: Тут возникают два связанных вопроса. Во-первых, каким образом мы вообще получаем знания? Во-вторых, хотим ли использовать для манипуляции ими символическую логику?

Мне кажется, что от символической логики не следует отказываться, и надеюсь, что кто-то найдет способ ее добавить. Самым крупным проектом такого рода был Сус Дуга Лената, начатый примерно в 1984 г. Этот закрытый проект оказался не очень эффективным. Сейчас, когда о машинном обучении известно намного больше, AI2 предпочитает делать вещи с открытым исходным кодом, в работе над которыми может принять участие сообщество.

М. Ф.: Когда, по вашим расчетам, может появиться сильный ИИ?

Г. М.: Понятия не имею. Есть множество причин, не дающих создать сильный ИИ здесь и сейчас. Можно назвать разве что доверительный интервал, как в статистике: период между 2030 г. и, в случае большой удачи, 2050 г., а в худшем случае — 2130 г.

Думаю, что в 1994 г., работая над книгой «Дорога в будущее», Билл Гейтс не осознавал, как сильно интернет изменит мир. ИИ хорошо финансируется, и, возможно, мы сможем продвинуться вперед.

М. Ф.: Итак, вы предполагаете, что прорыв может случиться или через 12, или через 112 лет.

Г. М.: В сильном ИИ особых достижений не наблюдается. Например, персональный помощник Siri, появившийся в 2010 г., не сильно отличается от созданной в 1966 г. программы ELIZA, которая пародировала диалог с психотерапевтом, реализуя технику активного слушания. До понимания естественного языка нам по-прежнему далеко. Но я все равно склонен к оптимистичным прогнозам.

М. Ф.: Да, ИИ больше не ограничен стенами университетов, он занял центральное место в бизнес-моделях крупных компаний, таких как Google и Facebook.

Г. М.: Сейчас на ИИ тратится намного больше денег, чем раньше, хотя в 1960-х гг., до начала так называемой «зимы ИИ» инвестиции в эту сферу также были достаточно большими. Но важно понимать, что хотя деньги — необходимое условие для продолжения исследований, они не гарантируют, что решение проблем будет найдено.

М. Ф.: Тогда давайте поговорим о более узкой технологии — беспилотных автомобилях. Когда, по вашему мнению, в произвольное место можно будет вызвать такси, которым управляет ИИ?

Г. М.: Как минимум через 10 лет.

М. Ф.: Практически такой же срок вы называли в прогнозе для сильного ИИ.

Г. М.: Дело в том, что вождение в мегаполисах связано с необходимостью постоянно обрабатывать непредсказуемые ситуации.

Для ИИ могут стать проблемой даже такие простые элементы, как защитные барьеры. Людей в таких случаях выручает здравый смысл. Современные беспилотные автомобили используют карты с высокой детализацией и технологию LIDAR, зрительная система человека работает хуже машинной, но люди лучше понимают, что происходит вокруг. Я не уверен, что простым увеличением количества воспринимаемых данных можно достичь точности, необходимой для вождения на Манхэттене. Можно получить точность 99,99 %, но даже этого будет недостаточно.

М. Ф.: Может, начинать нужно с автомобилей, которые ездят по заранее определенному маршруту?

Г. М.: Есть вероятность, что подобное скоро появится в городе Феникс. Осталось разработать маршруты без поворотов налево, чтобы как можно меньше приходилось пересекаться с людьми. По такому принципу уже работают монорельсовые дороги в аэропортах.

На беспилотный автомобиль влияет все — мокрый снег, слякоть, град, листопад, вещи, которые могут упасть с проезжающего мимо грузовика. Чем дальше мы продвигаемся в сторону системы, свободно взаимодействующей с окружающим миром, тем больше проявляется сложностей. У нас уже есть системы, которые приближаются к сильному ИИ по своей открытости. Именно поэтому мой прогноз в обоих случаях оказался практически одинаковым.

М. Ф.: Вы согласны с тем, что мы находимся на пороге новой промышленной революции, которая полностью изменит рынок труда?

Г. М.: Это действительно произойдет, хотя и медленнее, чем думают многие. Например, внедрение беспилотных автомобилей оказалось более сложным делом, чем предполагалось, поэтому водителям пока волноваться не о чем, а вот многочисленные сотрудники ресторанов быстрого питания и кассиры скоро могут

остаться без работы. Рано или поздно фундаментальные изменения неизбежно произойдут.

Проблемы из-за автоматизации начнутся к 2030 г. или, в лучшем случае, к 2070 г. Придется менять структуру общества, потому что в какой-то момент доступных рабочих мест станет меньше, чем трудоспособного населения.

Некоторые говорят, что после исчезновения большинства рабочих мест в сельскохозяйственном секторе появились новые рабочие места в промышленности, но мне этот аргумент кажется неубедительным. На получение первого работающего алгоритма для беспилотного автомобиля может потребоваться лет 50 работы и миллиарды долларов, но как только такие алгоритмы появятся, миллионы водителей потеряют работу в течение нескольких лет.

Новые профессии, конечно, появляются, но там не требуется большое количество людей. Например, предприниматель на YouTube может зарабатывать миллионы долларов, снимая видео. Но я сомневаюсь, что в этой нише смогут найти место потерявшие работу водители. В эпоху, когда небольшая группа может создать такой проект, как Instagram, трудно представить новую отрасль, которая предоставит миллионы рабочих мест.

М. Ф.: Примерно половина трудоустроенных людей сейчас занимается рутинной предсказуемой деятельностью, которая рано или поздно подвергнется автоматизации.

Г. М.: На самом деле все не так просто, как кажется на первый взгляд. В отличие от людей, ИИ-системы не понимают данных, представленных на естественном языке. Например, машина не может самостоятельно извлекать информацию из медицинских карт. Это рутинная и совсем несложная задача, но до ее автоматизации пока далеко.

М. Ф.: Можно ли решить проблему безработицы введением универсального базового дохода?

Г. М.: Реальной альтернативы универсальному базовому доходу я не вижу. И он неминуемо будет введен. Вопрос в том, удастся ли мирно договориться или начнутся массовые беспорядки.

М. Ф.: В США сейчас наблюдается настоящая эпидемия наркомании, которая вполне может быть связана с потерей работы из-за автоматизации на предприятиях и отчаянием, которое испытывают люди.

Г. М.: Возможно, это правда, но я не вижу однозначной связи между потерей работы и наркоманией. Более того, для многих наркотиком становятся смартфоны. Допускаю, что в итоге мы окажемся в мире, состоящем из виртуальной реальности и пособий.

М. Ф.: Но ИИ несет в себе и другие опасности. Например, Илон Маск активно говорил об экзистенциальных угрозах. Как вы думаете, у нас есть повод для беспокойства?

Г. М.: Беспокоиться следует о том, что люди могут использовать ИИ во зло. Именно в этом заключается реальная проблема, ведь ИИ все сильнее встраивается в нашу жизнь и становится все более уязвимым для хакерских атак. Нет повода переживать из-за того, что системы ИИ захотят съесть нас или превратить весь мир в скрепки. Я не утверждаю, что это невозможно, но реальных доказательств того, что подобное может случиться, нет. Зато очевидно, что машины становятся все более мощными, а мы пока толком не умеем решать вопросы информационной безопасности.

М. Ф.: А что вы думаете о проблеме выравнивания и рекурсивном цикле самосовершенствования?

Г. М.: Не скажу, что вероятность такого сценария равна нулю, но вряд ли подобное произойдет в ближайшее время. Недавно в сеть выложили видео о роботах, открывающих дверные ручки. Там рассказывается, на какой стадии находятся все эти разработки.

Систем ИИ, умеющих надежно ориентироваться в нашем мире, пока не существует. Нет и роботизированных систем, которые умеют себя улучшать. Так что проблема выравнивания не актуальна. Разумеется, имеет смысл инвестировать немного денег, чтобы кто-то искал решение на подобный случай. И заниматься насущными проблемами, такими как генерация и целенаправленная реклама недостоверной информации.

М. Ф.: Нужно ли следить за тем, чтобы цели интеллектуальной системы совпадали с нашими?

Г. М.: Разумеется, нужно. Но даже когда сильный ИИ будет создан, кто сказал, что у него будет желание вмешиваться в человеческие дела? Около 60 лет назад ИИ не мог выиграть партию в шашки, а в прошлом году решил куда более сложную задачу — выиграл в го. Можно сказать, что за 60 лет отвечающий за игры IQ вырос с 0 до 60. Если попытаться таким же способом оценить, насколько изменилась злобность машин, мы увидим, что никакой корреляции тут нет.

М. Ф.: Получается, что до появления сильного ИИ не может быть и речи о недовольстве машин.

Г. М.: Получается, что так. Отчасти это связано с системами мотивации. Например, у людей существует генетическая предрасположенность к совершению насильственных действий. Я сейчас о том, что мужчины склонны к насилию больше женщин.

М. Ф.: Получается, сильный ИИ должен быть женщиной?

Г. М.: Нужно просто сделать ИИ не склонным к насилию. Необходимы ограничения и правила, уменьшающие вероятность агрессивного поведения ИИ, направленного на человека. Это трудные и важные вопросы, куда менее очевидные, чем высказывания Маска, но они могут заставить многих людей задуматься.

М. Ф.: Но при этом Маск инвестирует в компанию OpenAI, что совсем неплохо. Эту работу все равно должен кто-то делать. Государственные инвестиции в вопросы будущего контроля ИИ было бы сложно понять, но почему бы этим не заниматься частным организациям?

Г. М.: Министерство обороны США действительно тратит деньги на подобные вещи, но должен быть портфель рисков. Меня больше беспокоят определенные виды биологического терроризма и кибернетические войны, чем разговоры о том, что ИИ может взять нас под контроль.

В подобных случаях нужно, во-первых, ответить на вопрос, можно ли говорить о том, что вероятность X больше 0. Ответ: однозначно да. Во-вторых, оценить вероятность рассматриваемого сценария по отношению к другим угрожающим сценариям. Тут я бы сказал, что это маловероятный сценарий.

М. Ф.: Как вы считаете, будет ли сильный ИИ обладать сознанием или станет зомби?

Г. М.: Скорее, второе. Я не считаю сознание обязательным условием. Это эпифеномен живых существ. В конце концов, мы можем столкнуться с чем-то, что ведет себя совсем как мы, но не обладает сознанием. Кроме того, пока нет способов определить наличие сознания. Откуда мне знать, обладаете ли сознанием вы?

М. Ф.: Можно предположить, что обладаю, потому что мы с вами принадлежим к одному виду.

Г. М.: На мой взгляд, это не аргумент. Что если сознание случайным образом распределено среди четверти населения? Что если за его наличие отвечает какой-то ген? Я, конечно, шучу, но факт остается фактом — объективных критериев, позволяющих установить наличие сознания, у нас нет.

М. Ф.: Получается, задача неразрешима.

Г. М.: Со временем кто-то, возможно, придумает решение, но пока по большей части исследуется та часть сознания, которую мы называем осознанностью. В какой момент центральная нервная система логически осознаёт, что определенная информация доступна?

Исследования показали, что если человек наблюдал что-то всего 100 миллисекунд, он не осознает, что видел это. Если наблюдение длилось половину секунды, человек практически уверенно осознает, что именно он видел. Базируясь на этих данных, уже можно создавать схему, указывающую, какие цепочки нейронов предоставляют информацию, доступную для осмысления и интерпретации. Именно это мы можем назвать осознанностью.

М. Ф.: А может получиться так, что сильный ИИ никогда не будет создан?

Г. М.: Нет. Он неизбежно появится, если человечество сохранится как вид. Нет причин не создавать его.

М. Ф.: Что вы думаете о международной гонке вооружений в сфере ИИ?

Г. М.: Китай открыто заявил, что ИИ у них одно из приоритетных направлений. США какое-то время никак на это не отвечали, что меня расстроило.

М. Ф.: Китай обладает многочисленными преимуществами: большей численностью населения и меньшей озабоченностью вопросами конфиденциальности, что позволяет получать больше данных.

Г. М.: Они дальновидны, понимают важность ИИ и инвестируют в него на государственном уровне.

М. Ф.: Должно ли правительство регулировать исследования в сфере ИИ?

Г. М.: Да, этой сфере требуется регулирование, но я не знаю, по какому принципу оно должно осуществляться. Например, мне совсем не нравится идея автономного оружия, но его прямой запрет, скорее всего, окажется наивной мерой, которая приведет к ситуации, когда кто-то обладает таким оружием, а кто-то безоружен.

М. Ф.: Вы верите, что ИИ окажет позитивное влияние на человечество?

Г. М.: Я на это надеюсь. Наибольшую пользу он мог бы принести здравоохранению. Но современные исследования и внедрение ИИ в основном связаны с рекламой.

Мне кажется, что огромному положительному потенциалу, которым обладает ИИ, уделяется недостаточно внимания, несмотря на то что внедрение ИИ будет сопровождаться исчезновением рабочих мест и социальными потрясениями. Сейчас мне не нравится, как используется ИИ и как мы его внедряем.

> *Я в восторге от того, что ИИ появился и начал менять наш мир, потому что не думала его застать, ведь преграды на пути к нему казались слишком сложными.*

Барбара Грош

ПРОФЕССОР ЕСТЕСТВЕННЫХ НАУК В ШКОЛЕ ИНЖЕНЕРНЫХ И ПРИКЛАДНЫХ НАУК ГАРВАРДА, ВНЕШТАТНЫЙ ПРЕПОДАВАТЕЛЬ НЕЗАВИСИМОГО ИНСТИТУТА ТЕОРЕТИЧЕСКИХ ИССЛЕДОВАНИЙ В САНТА-ФЕ

Благодаря Барбаре Грош сформировались основополагающие принципы обработки диалогов, необходимые для функционирования личных помощников, таких как Siri и Alexa. В 1993 г. она стала первой женщиной-президентом AAAI. Ее новаторские исследования в области обработки естественного языка и теории многоагентного взаимодействия внесли большой вклад в развитие ИИ. В настоящее время применяет разработанные модели для улучшения координации в сферах здравоохранения и образования.

Получила степень бакалавра математики в Корнеллском университете, а степень магистра и доктора computer science — в Калифорнийском университете в Беркли. Она является членом Национальной инженерной академии, Американского философского общества, Американской академии искусств и наук, а также AAAI и ACM. Обладательница премий ACM/AAAI Allen Newell Award, Award for Research Excellence от IJCAI и Lifetime Achievement Award от Ассоциации компьютерной лингвистики. Известна своим вкладом в развитие междисциплинарных учреждений и улучшение положения женщин в науке.

Мартин Форд: Как возник ваш интерес к ИИ?

Барбара Грош: Мой путь к ИИ — результат серии счастливых случайностей. В 18 лет я мечтала стать учителем математики. Потому что наш учитель в колледже верил в женщин-математиков и хвалил меня. Я поступила в колледж Корнеллского университета, где только что появился факультет computer science.

В то время в США получить степень бакалавра в области computer science было невозможно, но в Корнелле читали лекции по этой специальности. Я начала изучать методы численного анализа, затем поступила в магистратуру, а затем и в аспирантуру в Беркли.

Я работала в области, которая затем превратилась в computational science[1], а позже стала заниматься теоретической информатикой. Когда пришло время выбирать тему диссертации, Алан Кей предложил написать программу, которая будет читать детскую историю и пересказывать ее от лица одного из персонажей. Именно это вызвало у меня интерес к обработке естественного языка и привело в сферу ИИ.

М. Ф.: Алан Кей? Человек, который, работая в научно-исследовательском центре Xerox PARC, изобрел графический интерфейс пользователя? Именно его идеи Стив Джобс использовал при разработке компьютеров Macintosh.

Б. Г.: Да, верно, Алан был ключевым игроком в работе над Xerox PARC. На самом деле я работала с ним над созданием объектно-ориентированного языка программирования Smalltalk. Целью было создание системы, подходящей для обучения в детских садах и школах. Моя детская сказочная программа должна была

[1] Научное направление, связанное с построением математических моделей и разработкой численных методов для анализа и решения научных и инженерных задач с помощью компьютеров, например в физике, лингвистике, биохимии и т. д. — *Примеч. ред.*

быть написана на Smalltalk. Однако до того, как система Smalltalk была закончена, я поняла, что детские истории — это не просто истории, которые нужно читать и понимать. Они призваны прививать культуру, и задачу Алана мне будет очень тяжело выполнить.

В то время первая группа систем понимания речи также разрабатывалась в рамках проектов DARPA (Управления перспективных исследовательских проектов Министерства обороны США), и сотрудники SRI International, работающие над одной из них, сказали мне: “Если вы готовы рискнуть, работая над детскими историями, почему бы вам не поработать с нами над более предметным языком, направленным на диалог, но использующим речь, а не текст”. В результате я включилась в работу над системами, которые помогали людям в выполнении задач, и именно тогда начала проводить исследования в области искусственного интеллекта. Именно эта работа привела меня к открытию того, что диалог между людьми, работающими вместе над задачей, имеет структуру, которая зависит от структуры задачи, и что диалог — это гораздо больше, чем просто пары вопросов-ответов. Исходя из этого, я поняла, что как люди мы вообще никогда не говорим обособленными высказываниями, расставленными в нужной последовательности. Всегда существует структура, как, например, в журнальной или газетной статье, учебнике, и что мы можем смоделировать эту структуру. Это был мой первый крупный вклад в обработку естественного языка и искусственного интеллекта.

М. Ф.: Идея представить математически структуру диалога была прорывом в области естественного языка. Каким образом вы к ней пришли?

Б. Г.: Изначально перед нами стояла задача построить систему, умеющую вести естественный беглый диалог с человеком. Алана Кея тоже интересовало создание систем, которые будут адаптироваться к людям, а не наоборот.

В то время в лингвистике велись работы над синтаксисом и формальной семантикой, а в computer science — над алгоритмами синтаксического анализа. Уже была известна огромная роль контекста, но не было инструментов, математического описания и вычислительных конструкций для учета контекста в речевых системах.

Нам нужно было получить образцы реальных диалогов, которые ведут люди, совместно решающие некую задачу. Для этого двух человек, играющих роли эксперта и ученика, посадили в разные комнаты, чтобы исключить невербальную передачу информации, и попросили эксперта объяснить ученику, как выполнить некие действия. Проанализировав полученные диалоги, мы смогли понять их структуру и то, как она зависит от структуры задачи.

Позднее совместно с Кенди Сиднер мы написали статью *Attention, Intentions, and the Structure of Discourse* («Внимание, намерения и структура дискурса»), в которой рассказали, что диалоги имеют структуру, которая обусловлена самим языком, причинами вступления в диалог и целями каждого участника. Эта интенциональная структура представляет собой обобщение структуры задачи. Всеми этими аспектами управляет модель состояния внимания.

М. Ф.: Если сравнивать разработки «тогда» и «сейчас», что сильнее всего изменилось?

Б. Г.: Мы перешли от практически глухих речевых систем к системам, которые потрясающе обрабатывают речь. Улучшился анализ предложений и вычленение значений из них.

Но диалоговые системы, по сути, не работают. Они отлично справляются со всем, что попадает в рамки заданных сценариев, но реальные люди редко так разговаривают. Иногда ошибки, которые совершает система, создают серьезную этическую проблему.

Аналогичным образом обстоят дела со встроенными в смартфоны персональными помощниками. Например, если спросить, где находится ближайший травмпункт, вы получите его адрес, а вот в ответ на вопрос, куда обратиться с вывихнутой лодыжкой, система, скорее всего, просто откроет веб-страницу с информацией о способах лечения растяжений.

С этими проблемами сталкиваются и диалоговые системы, способные обучаться на данных. Летом 2017 г., когда Ассоциация по компьютерной лингвистике вручала мне награду, я обратилась к тем, кто работает над системами естественного языка на базе глубокого обучения, и сказала, что микроблоги Twitter не годятся в качестве примеров диалогов — нужны реальные данные.

М. Ф.: Но ведь способность отойти от сценария и справиться с непредсказуемой ситуацией это и есть настоящий интеллект. Именно здесь проходит граница между автоматом или роботом и человеком.

Б. Г.: Вы совершенно правы. Вспомните философскую идею коммуникативной импликатуры, разработанную в 1960-х гг. Полом Грайсом, Джоном Остином и Джоном Серлем. Например, говоря компьютеру: «Принтер неисправен», человек хочет, чтобы система предприняла какие-то действия для устранения неисправности, а не просто ответила: «Спасибо, факт зафиксирован». Но такое возможно только в случае, когда система может вычленить смысл из того, что было сказано.

Современные системы игнорируют интенциональную структуру диалога. Другие признаки интеллекта в системах на базе глубокого обучения в большинстве случаев тоже отсутствуют: они не могут использовать контрафактное мышление или рассуждать с точки зрения здравого смысла. Все эти вещи нужны для участия в свободном диалоге, когда на слова и действия обеих сторон не наложено никаких ограничений.

М. Ф.: Помню, как меня поразила победа суперкомпьютера Watson от IBM в телевикторине «Своя игра». Это действительно крупный прорыв или же существуют какие-то более передовые разработки?

Б. Г.: Я считаю Siri и Watson феноменальными достижениями техники. Они поменяли наш способ взаимодействия с компьютерами. Но до человеческих способностей им все еще далеко.

Когда в 2011 г. я впервые поговорила с Siri, мне понадобилось всего три вопроса, чтобы понять, что передо мной не человек. А когда допускает ошибки Watson, сразу становится ясно, где именно язык обрабатывается не так, как это делают люди.

Но мы действительно феноменально продвинулись. Я в восторге от того, что ИИ появился и начал менять наш мир, потому что не думала его застать, ведь преграды на пути к нему казались слишком сложными.

М. Ф.: Вы действительно не верили, что ИИ появится при вашей жизни?

Б. Г.: Тогда, в 1970-х гг.? Нет, не верила.

М. Ф.: Watson удивил меня умением понимать каламбуры, шутки и сложные языковые конструкции.

Б. Г.: Если заглянуть внутрь этих систем, сразу станут заметными их ограничения. Для сферы ИИ, да и, честно говоря, для всего мира было бы полезнее, если бы мы не стремились заменить людей или создать сильный ИИ, а попытались понять, для чего лучше подходят способности компьютера и как мы можем взаимодополнять друг друга.

М. Ф.: Способность говорить не по сценарию напрямую связана с тестом Тьюринга. Я знаю, что вы внесли свой вклад в эту область. Как вы думаете, зачем Тьюринг придумал этот тест?

Б. Г.: Я напомню, что Тьюринг предложил свой тест в 1950 г., после появления новых вычислительных машин. Конечно, по сравнению с тем, что предлагают современные смартфоны, возможности этих систем были невелики, но уже тогда многие задавались вопросом, могут ли эти машины думать как человек. Если помните, Тьюринг приравнивал термины «интеллект» и «мышление». Он поставил интересный философский вопрос — могут ли машины демонстрировать определенный тип поведения? В то время психология базировалась на бихевиоризме, поэтому его тест представлял собой испытание эксплуатационных качеств, без учета того, что происходит внутри.

Тест Тьюринга нельзя считать хорошим тестом на интеллект. Честно говоря, я бы его, скорее всего, провалила. Не показывает он и того, в какую сторону нужно развивать ИИ. Тьюринг был удивительно умным человеком, и, мне кажется, он бы предложил другой вариант теста, если бы жил сейчас.

М. Ф.: Я знаю, что и вы предложили усовершенствование или даже замену тесту Тьюринга.

Б. Г.: Сейчас мы знаем, что развитие человеческого интеллекта, как и языковой потенциал, зависит от социального взаимодействия. Кроме того, во многих ситуациях люди предпочитают действовать совместно. Поэтому я предположила, что нужно стремиться создать систему, которая была бы хорошим партнером и работала бы в коллективе так хорошо, что люди бы даже не задумывались о ее природе. Дело ведь не в том, чтобы одурачить людей, заставив их думать, что ноутбук, робот или телефон — это такой же человек. Нужно сделать так, чтобы у людей не возникало вопроса «почему он поступил именно так?». А такой вопрос непременно возникает, когда совершается ошибка, не характерная для человека.

Поставленная таким образом цель имеет несколько преимуществ по сравнению с тестом Тьюринга: к ней можно двигаться постепенно, выбирая оптимальное направление, и получить взаимодо-

полнение способностей человека и компьютера. Мой тест высоко оценили в Эдинбурге на мероприятии, посвященном 100-летию со дня рождения Тьюринга.

М. Ф.: Я всегда полагал, что как только мы действительно получим машинный интеллект, это попросту будет очевидно всем. Ведь в процессе общения мы каким-то образом без предварительных тестов понимаем, умный перед нами человек или не очень.

Б. Г.: Это хорошее замечание. Если вспомнить мой пример с двумя вопросами «где ближайшая больница?» и «где искать помощь в случае сердечного приступа?», то человек (не приезжий), который считается умным, даст ответ в обоих случаях. Мой вариант теста не ограничен во времени и изначально задумывался как пролонгированный. Впрочем, тест Тьюринга также не предполагал временных ограничений, но об этом часто забывают, в частности, на соревнованиях по ИИ.

М. Ф.: Это же абсурдно. Чтобы продемонстрировать интеллект, нужен неопределенно долгий период времени. Мне вспоминается премия Лебнера, которая ежегодно присуждается программам по результатам прохождения ими ограниченного определенным временем теста Тьюринга.

Б. Г.: Именно так. Кроме того, это объясняет одну вещь, с которой мы столкнулись, когда только начинали заниматься обработкой естественного языка. Оказалось, что в случае фиксированной задачи с фиксированным набором ответов на нее, которую нужно решить за фиксированное время, приемы, легко дающие нужный результат, предпочитаются настоящей интеллектуальной обработке данных. Фактически система проектируется таким образом, чтобы она могла пройти тест!

М. Ф.: Расскажите о второй области, в которой вы работали, — многоагентных системах.

Б. Г.: Интенциональную модель дискурса, о которой я упоминала ранее, мы с Кенди Сиднер сначала попытались построить на базе работы наших коллег, которые использовали для формализации теории речевых актов разработанные для роботов ИИ-модели планирования. Но в контексте диалога описанные техники давали неадекватные результаты. Мы поняли, что командную работу или совместную деятельность нельзя представлять как сумму индивидуальных планов.

В то время исследователи ИИ часто использовали примеры, в которых фигурировала сборка из игрушечных блоков. Я представила, как два ребенка строят башню, при этом у одного есть набор синих блоков, а у другого — красных. В случае суммы индивидуальных планов мы получим абсурдную ситуацию: у ребенка с синими блоками есть план, как расположить их в пространстве, который четко совпадает с тем, где по плану ребенка с красными блоками остаются пустые места.

Мы с Сиднер поняли, что для системы, состоящей из нескольких участников, нужен новый способ обдумывания планов и представления их в компьютерной системе. Именно с этого началась разработка многоагентных систем.

В 1980-х гг. в этой области в основном рассматривались ситуации с несколькими роботами или компьютерными агентами и ставились вопросы конкуренции и координации.

М. Ф.: Под компьютерным агентом вы понимаете программу, которая запускает и выполняет какое-то действие?

Б. Г.: В общем случае это система, способная действовать автономно. Первоначально большинство компьютерных агентов были роботами, но постепенно в исследованиях ИИ стали появляться и программные агенты. Сегодня они выполняют поиск, торгуются на аукционах и решают множество других задач.

Например, Джефф Розенхейм рассматривал ситуацию с роботами-курьерами, пытаясь понять, как скажется на эффективности их работы возможность обмена пакетами. Он задавался вопросом, будут ли роботы всегда говорить правду о стоящих перед ними задачах, ведь в некоторых случаях, солгав, можно добиться лучшего результата.

Сейчас в области многоагентных систем решаются задачи, связанные со стратегическим мышлением и командной работой. Приятно, что в последнее время многие начали рассматривать вопросы взаимодействия компьютерных агентов с людьми, а не только с другими компьютерными агентами.

М. Ф.: Я правильно понимаю, что именно эта работа привела к компьютерной модели сотрудничества?

Б. Г.: Да, первая компьютерная модель сотрудничества появилась как следствие работы с многоагентными системами.

Нужно было ответить на вопрос, что такое сотрудничество. Люди делят общую задачу на набор более мелких задач, которые делегируются разным участникам. Каждый обязуется выполнить свою часть, не забывая об общем деле.

Вместе с Сарит Краус мы формализовали все эти вещи в рамках компьютерной модели, что породило множество новых вопросов: как распределить задачи, что произойдет в нестандартной ситуации и какие обязательства перед командой имеет каждый участник.

В 2011–2012 гг. во время академического отпуска я решила посмотреть, могут ли мои разработки на тему сотрудничества изменить мир. С этого началась моя работа в сфере здравоохранения. Вместе с педиатром из Стэнфорда Ли Сандерсом мы стали искать новые методы координации. Особое внимание мы уделили детям со сложными заболеваниями, которые посещают множество врачей. Нужно было создать систему, которая

поможет этим врачам обмениваться информацией и успешно координировать свои действия.

М. Ф.: На мой взгляд, здравоохранение — одна из наиболее перспективных областей для применения ИИ.

Б. Г.: Правильно. Как и в сфере образования, здесь важно сосредоточиться на создании систем, дополняющих, а не замещающих людей.

М. Ф.: Как вы относитесь к тому, что сейчас везде говорят о глубоком обучении? Мне кажется, что в результате у людей складывается впечатление, что глубокое обучение — это синоним ИИ.

Б. Г.: Если брать философские смыслы, глубокое обучение не позволяет получить систему, которая мыслит глубже, чем системы, обученные другими способами. Оно хорошо функционирует благодаря большей математической гибкости и подходит для выполнения задач по принципу сквозной обработки: на вход поступает сигнал, на выходе мы получаем ответ. Но этот ответ зависит от того, на каких именно данных система обучена.

М. Ф.: Не приведет ли признание всех этих ограничений к движению против глубокого обучения?

Б. Г.: Как человек, уже переживавший «зимы ИИ», я чувствую страх и надежду. Чтобы избежать «зимы», нужно четко определить, какое место в создании ИИ занимает глубокое обучение.

М. Ф.: Думаю, нужно работать над системами, умеющими обучаться на меньших количествах данных.

Б. Г.: При этом вопрос не столько в количестве данных, сколько в их разнообразии. Сейчас мы имеем системы, качество работы которых зависит от того, к какой группе принадлежит пользователь.

М. Ф.: Вы сейчас имеете в виду систему распознавания лиц, которая испытывает трудности при работе с изображениями представителей негроидной расы?

Б. Г.: Это простейший пример, который сразу приходит на ум. Но проблема смещения данных присутствует во всех областях. Скажем, при разработке нового лекарства исследования в основном проводятся на молодых людях. В результате сложно определить корректную дозировку для пожилых.

М. Ф.: Систематическая ошибка в данных, обусловленная решениями, которые принимались в процессе исследований, имеет место не только в сфере ИИ.

Б. Г.: Именно так. Компьютерная система может «прочесть все документы», извлечь из них нужную информацию и провести статистический анализ. Но если в большинстве статей описываются результаты опытов на самцах мыши, система не сможет прийти к обобщенным выводам.

Кроме того, существуют ограничения, связанные с правовым полем, с охраной данных и с законностью их использования. Об этом тоже нужно думать на стадии построения ИИ-систем.

М. Ф.: В отличие от вас, многие специалисты заинтересованы в создании независимых машин.

Б. Г.: Мне кажется, они читают слишком много научно-фантастических романов!

М. Ф.: Тогда сильный ИИ — тоже научная фантастика? Что, с вашей точки зрения, мешает его создать?

Б. Г.: В конце 1970-х гг., когда я заканчивала работу над диссертацией, мой однокурсник сказал: «Хорошо, мы не ставим перед собой цель заработать денег, ведь за нашу работу никогда не будут много платить». Я часто вспоминаю этот прогноз и понимаю,

что хрустального шара, позволяющего заглянуть в будущее, не существует.

С моей точки зрения, не нужно стремиться к сильному ИИ, потому что его появление влечет за собой такие вещи, как массовая безработица и выход роботов из-под контроля. Эти вопросы отвлекают специалистов от актуальных этических проблем с уже существующими ИИ-системами.

Разумеется, люди в течение многих сотен лет задумывались над тем, можно ли создать нечто такое же умное, как человек. Вспомните хотя бы пражского Голема или Франкенштейна. Невозможно запретить людям фантазировать, главное не тратить на это наши ресурсы и интеллект.

М. Ф.: А все-таки, какие препятствия стоят на пути к созданию сильного ИИ?

Б. Г.: Выше я уже упоминала одно из главных препятствий. Нужно получить широкий спектр данных, но это означает, что придется, как Большой Брат, следить за поведением огромного количества людей. Пока непонятно, как перейти к общему интеллекту, умеющему гибко перемещаться из одной области в другую и думать не только о настоящем, но и о будущем.

М. Ф.: Но для сокращения количества рабочих мест специализированных ИИ-систем вполне достаточно. Вы считаете это поводом для беспокойства?

Б. Г.: Да, у нас есть повод для беспокойства, но это более широкая технологическая проблема, чем внедрение ИИ. За ее возникновение ответственны как создатели технологий, так и представители деловых кругов. Помимо вопросов сокращения сотрудников, какой бы интеллектуальной ни была машина, она будет проигрывать квалифицированному работнику, что приведет к падению качества обслуживания клиентов. Именно

поэтому я так много говорю о необходимости систем, которые дополняют людей.

М. Ф.: Я считаю, что в данном случае речь идет о столкновении технологий и капитализма.

Б. Г.: Вы совершенно правы!

М. Ф.: Исторически стремление заработать, сокращая расходы, было положительным моментом. Но сейчас нужно что-то менять, пока не началось беспрецедентное вытеснение людей с рабочих мест.

Б. Г.: С этим я полностью согласна. Недавно я говорила об этом в Американской академии наук и искусств, выделяя два ключевых момента. Во-первых, вопрос не в том, какие системы мы можем построить, а в том — какие должны. У создателей технологии всегда есть выбор, даже в капиталистической системе. Во-вторых, в курс computer science следует добавить этику, чтобы студенты думали не только об эффективности и элегантности кода, но и о моральной стороне разработок. На этой встрече я приводила в пример компанию Volvo, конкурентным преимуществом которой стали безопасные автомобили. Нужно сделать так, чтобы конкурентным преимуществом стали системы, которые хорошо работают с людьми. Но для этого инженеры должны сотрудничать с социологами и специалистами по этике.

М. Ф.: Какие еще опасности, на ваш взгляд, в ближайшей перспективе и в будущем несет в себе ИИ?

Б. Г.: Лично у меня есть множество вопросов по поводу возможностей, которые дает ИИ, доступных ему методов и вариантов применения этих методов, а также по поводу дизайна ИИ-систем.

У нас всегда есть выбор. Даже когда речь заходит об оружии. Был выбор и у Илона Маска. Он мог сказать, что в автомобили Tesla встроена система помощи водителю, а не автопилот. Тем более что

полноценной машины с автопилотом на самом деле не существует. Люди покупаются на рекламу, верят, что автопилот работает, а затем попадают в аварию.

Мы можем выбирать, что мы поместим в систему, какие требования к ней предъявить, как ее тестировать, проверять и настраивать. Только от нашего выбора зависит, сможем ли мы избежать будущей катастрофы.

В состав проектных групп нужно включать людей, которые будут рассматривать возможные последствия создания различных систем. Я имею в виду юридические и этические последствия и прочие побочные эффекты. Пришло время прекратить развитие технологий по принципу «интересно, смогу ли я создать что-то, что делает вот такую вещь». Нужно думать, как в долгосрочной перспективе то, что мы создаем, повлияет на нашу жизнь. Это социальная проблема.

Поэтому я вела курс «Интеллектуальные системы: проектирование и этические проблемы», а сейчас вместе с коллегами из Гарварда работаю над программой Embedded EthiCS, позволяющей интегрировать преподавание этики в каждый курс computer science.

М. Ф.: Как вы считаете, не слишком ли много внимания уделяется экзистенциальным угрозам? Нужно ли относиться серьезно к проблемам, которые могут появиться только в отдаленном будущем?

Б. Г.: Уже сейчас можно легко создать дрон, несущий на борту оружие. Но я не разделяю экстремального подхода некоторых людей, которые считают, что нужно приостановить все разработки ИИ.

М. Ф.: Вы часто говорите, что у нас есть выбор. Но принимаемые решения в значительной степени обусловлены рыночными стимулами. Кто в итоге должен делать выбор — проектировщики, инженеры, коммерсанты или государство?

Б. Г.: Вспомните самообучающегося бота Tay от Microsoft, который под влиянием пользователей всего за один день стал расистом. Сфера ИИ нуждается в регулировании с привлечением социологов и специалистов по этике.

М. Ф.: Но не поставит ли это США или западные страны в целом в невыгодное положение в конкурентной гонке с Китаем?

Б. Г.: Если мы остановим все исследования и разработки в сфере ИИ или строго ограничим их, то действительно окажемся в невыгодном положении. Ключевой вопрос в том, что произойдет, если мы не создадим оружия, управляемого ИИ, а у противников оно будет. Но это слишком большая и серьезная тема.

М. Ф.: В заключение я хочу спросить, какой совет вы можете дать тем, кто хочет работать в сфере ИИ? Каким был ваш опыт?

Б. Г.: Прежде всего эта область затрагивает одни из самых интересных вопросов в мире. Совокупность вопросов, которые поднимает ИИ, всегда требовала сочетания аналитического и математического мышления, умения думать о людях и понимать их поведение, инженерного мышления. Вы получаете возможность исследовать всевозможные способы мышления и всевозможные дизайнерские решения.

Я уверена, что другие люди считают свои области самыми захватывающими, но сейчас ИИ еще более захватывающая область, потому что у нас есть гораздо более мощные инструменты: просто посмотрите на наши вычислительные мощности. Когда я начинала работать, одна из моих коллег вязала свитер в ожидании, пока закончится обработка кода и на экране снова появится приглашение на ввод!

Как и все работающие в этой сфере, я уверена, что очень важно привлекать к разработкам ИИ-систем как можно больше людей. Я имею в виду не только женщин и мужчин, а людей разных культур, разных рас, потому что именно они будут использовать

эти системы. Иначе возникают две опасности. Во-первых, такие системы, подойдут только определенным группам населения, а во-вторых, возникнет недружелюбный рабочий климат. Мы все должны работать вместе.

Я начинала, когда в этой сфере практически не было женщин, так что все зависело от отношения коллег-мужчин. Могу вспомнить как фантастически прекрасные, так и ужасные моменты. И считаю, что каждый университет и каждая компания, в которой есть группа, занимающаяся технологиями, должны поощрять и привлекать в эту сферу женщин и представителей меньшинств, потому что сейчас мы уже точно знаем, что чем многообразнее команда проектировщиков, тем лучше будет конечный результат.

> *Сейчас прогресс замедлен из-за концентрации на глубоком обучении и связанных с ним непрозрачных структурах.*

Джуда Перл

ПРОФЕССОР COMPUTER SCIENCE И СТАТИСТИКИ, ДИРЕКТОР ЛАБОРАТОРИИ КОГНИТИВНЫХ СИСТЕМ В UCLA

Джуда Перл — автор математического аппарата байесовских сетей, создатель математической и алгоритмической базы вероятностного вывода, автор алгоритма распространения доверия для графических вероятностных моделей.

Будучи выпускником Израильского технологического института, в 1960 г. приехал в США и через год получил степень магистра электротехники в Нью-Аркском инженерном колледже, ныне Технологическом институте Нью-Джерси. Степень магистра физики получил в Ратгертском университете, а докторскую степень по электротехнике — в Бруклинском политехническом институте. Занимался исследованиями в Лабораториях Давида Сарнова в Принстоне и компании Electronic Memories, Inc.

В 1969 г. начал работать в UCLA, где до сих пор возглавляет созданную им лабораторию когнитивных систем. Внес большой вклад в ИИ, теорию человеческого мышления и философию науки. Автор более 450 научных работ и трех знаковых книг:

Heuristics («Эвристика», 1984 г.), Probabilistic Reasoning («Вероятностные рассуждения», 1988 г.) и Causality: models, reasoning, and inference («Причинность: модели, рассуждения и вывод», 2000 и 2009 гг.).

Член Национальной академии наук, Национальной инженерной академии и член-основатель AAAI. Обладатель многочисленных научных наград, в том числе премии Тьюринга за «фундаментальный вклад в ИИ посредством разработки исчисления для проведения вероятностных и причинно-следственных рассуждений», премии Румельхарта за «вклад в теоретические основы человеческого познания», премии Харви за «основополагающие работы, оказавшие влияние на множество аспектов современной жизни», премии Лакатоса за «лучшую книгу в области философии науки», медали Аллена Ньюэлла за «плодотворный вклад в философию, психологию, медицину, статистику, математическую экономику, эпидемиологию и общественные науки» и медали Бенджамина Франклина в области computer science и когнитивных наук.

Мартин Форд: Как началась ваша карьера в computer science и ИИ?

Джуда Перл: Я вырос в Бней-Браке. Моей любознательности способствовало воспитание в рамках израильского общества и полученное мной уникальное образование. Тогда в старших классах школы и в колледжах преподавали первоклассные ученые, приехавшие из Германии в 1930-х гг. У них не было возможности найти работу в академических кругах, поэтому им пришлось работать в школах. Они понимали, что уже никогда не вернутся к занятиям наукой, и видели в нас воплощение своей мечты. Так что моему поколению крупно повезло. Я никогда не был лучшим учеником, но всегда увлекался тем, что нам преподавали. Информация нам преподносилась в хронологическом порядке, причем повышенное внимание уделялось людям,

которые стояли за каждым изобретением или теоремой. Это позволяло увидеть, что наука — это не просто набор фактов, а непрерывная борьба человека с окружающей его неопределенностью.

В армии я присоединился к кибуцу и собирался провести там всю жизнь, но умные люди сказали, что я буду счастливее, если стану развивать математические способности. В результате в 1956 г. я поступил в Израильский технологический институт и начал изучать электронику. Любимой специальности у меня не было, но мне нравились логический синтез и электромагнитная теория. В 1960 г. я получил степень бакалавра, женился и поехал в США — хотел получить докторскую степень и вернуться.

М. Ф.: Вы планировали вернуться в Израиль?

Дж. П.: Разумеется. Я поступил в Бруклинский политехнический институт, который в то время был одной из лучших школ радиорелейной связи. Но оплачивать обучение я не мог, поэтому работал в исследовательском центре Давида Сарнова в Лаборатории RCA в Принстоне. Там я входил в группу доктора Яна Райхмана, которая занималась запоминающими устройствами. Мы искали физические механизмы, пригодные для компьютерной памяти. Память на магнитных сердечниках работала слишком медленно, была громоздкой и требовала кропотливого ручного труда при производстве.

Когда стало понятно, что дни этой технологии сочтены, все — IBM, Лаборатории Белла и RCA — начали искать новые механизмы хранения цифровой информации. В то время из-за скорости и простоты подготовки были популярны сверхпроводниковые запоминающие устройства, хотя им и требовалось охлаждение до температуры жидкого гелия. Исследуя токи, циркулирующие в сверхпроводниках, я обнаружил несколько интересных явлений. В мою честь был даже назван пирловский вихрь — турбулентный поток в сверхпроводящих пленках. Это было потрясающее время.

В начале 1960-х гг. всех воодушевляли потенциальные возможности компьютеров. Никто не сомневался, что в конечном итоге компьютеры научатся выполнять большинство интеллектуальных задач лучше людей. Способы выполнения этих задач искали все, даже специалисты по аппаратному обеспечению. Мы постоянно думали, как создать ассоциативные воспоминания, как научить машины восприятию, распознаванию объектов, кодированию визуальных сцен. Руководство RCA всячески поощряло наши инициативы. Раз в неделю к нам заходил Райхман и спрашивал, что у нас можно запатентовать.

Разумеется, после появления полупроводников все работы по сверхпроводимости прекратились, хотя в то время мы не верили в успех новой технологии. Которая к тому же имела уязвимость — в случае разряда батареи память стиралась. Но решение было найдено, и полупроводниковые технологии уничтожили всех конкурентов. Я потерял работу в компании Electronic Memories. После этого я и попал в академические круги, где исполнил свою давнюю мечту — занялся распознаванием объектов и кодированием изображений.

М. Ф.: После увольнения из Electronic Memories вы сразу поступили в UCLA?

Дж. П.: Сначала я пытался поступить в Университет Южной Калифорнии, но меня подвела самоуверенность. Я претендовал на место преподавателя программирования, хотя нужного опыта у меня не было. Декан просто выгнал меня из офиса. В итоге я попал в UCLA, где получил возможность заниматься тем, чем хотел. Постепенно от распознавания объектов, кодирования изображений и теории принятия решений я перешел к ИИ. В первые годы в этой сфере занимались игровыми программами, которые, с моей точки зрения, могли послужить основой для программного моделирования человеческой интуиции. Решение этой задачи до сих пор остается мечтой моей жизни.

Именно интуиция позволяет оценивать эффективность ходов в игре. В этом между способностями машин и людей-экспертов был большой разрыв. Я проанализировал эвристический поиск и нашел способ его автоматизации, который применяется и сейчас. Кажется, я был первым, кто показал оптимальность альфа-бета-отсечения и других математических приемов. Они собраны в книге «Эвристика». Затем на сцену вышли экспертные системы, и люди пришли в восторг от моделей различных видов эвристической деятельности, таких как интуиция врача или геолога. Идея заключалась в том, чтобы заменить специалистов-людей или помочь им.

М. Ф.: Если я правильно помню, экспертные системы по большей части базируются на множестве правил? «Если верно это, то делай то» и т. п.

Дж. П.: Именно так. Но мы смоделировали не действия врача, а течение болезней, указав их симптомы. Появился принцип диагностики, который работает также для поиска полезных ископаемых, обнаружения неисправностей и для любой другой профессиональной деятельности.

М. Ф.: Основой этого послужили ваши работы по эвристике или вы уже говорите о байесовских сетях?

Дж. П.: Нет, эвристикой я прекратил заниматься в 1983 г. и начал работать над байесовскими сетями и управлением неопределенностью. Варианты управления неопределенностью тех лет не соответствовали требованиям теории вероятностей и теории принятия решений, а я хотел сделать все правильно и эффективно.

М. Ф.: Расскажите о своей работе над байесовскими сетями.

Дж. П.: Во-первых, следует упомянуть, в каких условиях мы тогда работали. Это было постоянное противостояние между халтур-

щиками и аккуратистами. Первые всегда преследовали единственную цель: построить работающую систему. Без гарантий, без проверок на соответствие теориям. Вторые хотели понять, почему система работает, и гарантировать эффективность ее функционирования.

М. Ф.: Вы говорите о противостоянии людей с разными позициями?

Дж. П.: Да. Сегодня в сообществе машинного обучения мы наблюдаем аналогичную ситуацию. Причем зачастую управляющие рычаги находятся в руках халтурщиков за счет связей со спонсорами. Представители индустрии зачастую недальновидны и требуют быстрых успехов, что смещает акценты исследований. Во время работы над байесовскими сетями я был среди немногих, кто выступал за то, чтобы делать все по правилам теории вероятностей. К сожалению, при таком подходе требовались экспоненциальные время и память. Таких ресурсов не было.

Меня вдохновила работа когнитивного психолога Дэвида Румельхарта, который исследовал, каким образом детям удается быстро и правильно читать тексты. Он предлагал создать многослойную систему, в которой с уровня пикселов шел переход на семантический уровень, затем — на уровень предложений и грамматики, при этом происходило установление связи и передача сообщений. Каждый уровень выполнял только свою задачу и не знал, что делают остальные, чтобы получить корректный ответ, в котором, например, слово car отличается от слова cat.

Попытки смоделировать такую архитектуру в теории вероятностей не давали хорошего результата, пока я не обнаружил, что свойство сходимости появляется при древовидной структуре модулей. Сообщения можно передавать асинхронно, и в конечном итоге система даст корректный ответ. Затем мы перешли к полидеревьям, и, в конце концов, в 1995 г. я опубликовал статью о байесовских сетях.

Программировать эту архитектуру оказалось на удивление легко. Не требовалось управляющей программы для наблюдения за всеми элементами. Достаточно было указать, что будет происходить с переменной, когда она решит обновить информацию. Затем эта переменная отправляла сообщения соседям, которые в свою очередь отправляли сообщения своим соседям и т. д., после чего система давала корректный ответ.

Байесовские сети были приняты благодаря простоте программирования. Более того, они позволили запрограммировать зависимость между симптомами и болезнью и вычислить вероятность заболевания по наличию или отсутствию симптомов. При этом пользователю было понятно, почему система дает тот или иной результат и как ее модифицировать в случае изменившейся среды. Преимуществом была и модульность, характерная для способов, которые работают в природе.

В то время мы не понимали значение модульности. Оказалось, что она обеспечивается причинностью. Убрав причинно-следственные связи, мы теряем модульность, и вместе с ней прозрачность, возможность менять конфигурацию и другие приятные функции. К 1988 г., когда вышла моя книга о байесовских сетях, я уже хотел перейти к следующему шагу — моделированию причинности.

М. Ф.: Мы то и дело слышим, что «взаимосвязь и причинно-следственные отношения — это разные вещи» и что данные не дают информации о причинно-следственных связях. Правильно ли я понимаю, что байесовские сети также не позволяют ее получить?

Дж. П.: Нет. Байесовские сети работают в разных режимах в зависимости от того, как сконструированы.

М. Ф.: Идея этих сетей в обновлении вероятностей на основе новых данных ради получения более точной оценки. Вы нашли, как эффективно сделать это для большого числа вероятностей.

Дж. П.: Теорема Байеса применяется давно, сложность состояла в том, чтобы найти эффективный способ ее применения. Я считал, что для машинного обучения это просто необходимо. Можно получить данные и с помощью теоремы Байеса обновить систему для повышения ее производительности и улучшения параметров. Но это вероятностная, а не причинно-следственная схема, поэтому она имеет ограничения.

М. Ф.: Но она часто используется в системах распознавания речи и на различных устройствах.

Дж. П.: Мне доводилось слышать, что байесовские сети используются во всех сотовых телефонах для подавления помех при передаче, как и алгоритм распространения доверия. Именно так мы назвали алгоритм передачи сообщений. Якобы они есть в Siri, но проверить это невозможно, так как Apple держит свои разработки в секрете.

Байесовский вывод — один из основных компонентов машинного обучения, но постепенно произошел переход от байесовских сетей к менее прозрачному глубокому обучению. Система теперь самостоятельно регулирует параметры, а мы понятия не имеем, что за функция соединяет вход и выход. Работая с байесовскими сетями, мы не до конца осознавали важность модульности. При моделировании диагностической системы туда закладывается причинно-следственная связь между симптомами и заболеваниями. Но при этом возникает вопрос: как выглядит тот ингредиент, который мы называем «причинно-следственными связями»? Где он находится, и как мы его обрабатываем?

М. Ф.: Байесовские сети стали популярными в computer science благодаря вашей книге. Но вы еще до ее выхода хотели перейти к причинному анализу?

Дж. П.: Именно причинно-следственные связи привели меня к интуитивному озарению, которое породило байесовские сети.

Теоретически можно обойтись без причинно-следственной связи. Все функции байесовской сети можно описать, используя чисто вероятностную терминологию. Но на практике оказалось, что структурирование сети с учетом причинности все сильно упрощает. Хотя мы и не понимали почему.

Теперь известно, что нужны были такие функциональные характеристики, как модульность, возможность перестройки структуры, переноса и многое другое. Все эти вещи обеспечиваются именно причинностью. Но при более детальном рассмотрении оказалось, что фраза «корреляция не подразумевает причинности» намного глубже, чем мы думали. Чтобы получить причинно-обусловленные выводы, нужны причинно-обусловленные предположения, которые невозможно взять из данных и вообще непонятно, как их выразить.

У нас не было языка, на котором можно было бы передавать простые вещи типа «грязь не вызывает дождя» или «петух не причина восхода солнца». Это нельзя записать математически, а значит, невозможно объединить с данными.

Словом, требовался новый язык. Для меня это стало шоком, потому что я вырос на статистике и считал, что эта наука позволяет все — индукцию, дедукцию, абдукцию и обновление модели. И внезапно оказалось, что язык статистики не может нам помочь. В computer science принято создавать языки под конкретные потребности. Но какой язык был нужен в нашем случае? Статистика использует язык средних значений, проверки гипотез, обобщения данных и визуализации их с разных точек зрения. И к этому всему нужно было добавить язык причин и следствий.

М. Ф.: И вы изобрели технический язык или диаграммы для описания причинно-следственной связи?

Дж. П.: Это не мое изобретение. Генетик Сьюалл Райт в 1920 г. первым проиллюстрировал причинно-следственные связи с по-

мощью стрелок и узлов и доказал, что такую информацию статистики не смогут получить из регрессии, ассоциации или корреляции.

Я преобразовал эти иллюстрации в средство кодирования научных знаний, позволяющее программировать машины на выяснение причинно-следственных связей в различных областях: от медицины и образования до причин глобального потепления — то есть в таких, где важны причины явлений, их механизмы и контроль.

Этим я занимался последние 30 лет. В 2000 г. вышла моя книга «Причинность: модели, рассуждения и вывод», а в 2009 г. — ее второе издание. В 2015 г. в соавторстве с Даной Маккензи я написал книгу *The Book of Why* («Книга о вопросе “Почему”»), в которой принцип причинности объяснялся широкой аудитории. В ней я рассматриваю историю с точки зрения причинности, рассказывая о том, какие концептуальные открытия изменили наше мышление.

М. Ф.: На мой взгляд, причинно-следственные модели стали важны в социальных и естественных науках именно благодаря вашей книге. Недавно я читал статью, в которой эти модели применялись в квантовой механике.

Дж. П.: Я тоже ее читал и собираюсь перечитать, потому что не понял явления, которые там описывались.

М. Ф.: Из «Книги о вопросе “Почему”» я понял, что естествоиспытатели и социологи действительно начали руководствоваться принципом причинности, а область ИИ в этом отношении отстает. Вы считаете, что для развития этой области нужно сосредоточиться на причинно-следственных связях?

Дж. П.: Да. В машинном обучении сейчас господствуют статистики, и считается, что из данных можно научиться всему. Но проис-

ходящее имеет четкие теоретические ограничения. Невозможно сделать гипотетические построения и продумать действия, с которыми вы раньше не сталкивались. Воображение — это высокий когнитивный уровень, позволяющий рассуждать о том, как бы выглядел мир, если бы события сложились по-другому.

Способность воображать сценарии, не чувствуя дискомфорта, позволяет людям создавать новое и исправлять старое, брать на себя ответственность, сожалеть и иметь свободную волю. Все это часть нашего умения генерировать миры, которые могли бы существовать. При этом есть правила создания правдоподобных гипотетических ситуаций, поняв которые мы сможем передать их машине. У меня есть несколько идей по этому поводу, ведь мы уже знаем, что нужно для интерпретации гипотетических ситуаций и понимания причин и следствий.

Это мини-шаги к сильному ИИ, но они могут многому нас научить. Я пытаюсь донести это до сообщества. Глубокое обучение — всего лишь шаг к сильному ИИ. Нужно только найти способ обойти барьеры, мешающие добавить причинно-следственные связи.

М. Ф.: Получается, человеческий разум обладает встроенным механизмом, позволяющим создавать каузальные модели, а не просто учиться на данных.

Дж. П.: И не только создавать. Ведь даже если модель создает кто-то со стороны, нужен механизм, позволяющий ее использовать.

М. Ф.: Получается, что каузальная модель — это просто гипотеза. У каждого человека может быть своя модель, и где-то в мозге некий механизм непрерывно генерирует такие модели, что дает возможность делать умозаключения на основе данных.

Дж. П.: Людям нужно создавать такие модели, модифицировать их и искажать, когда в этом возникает необходимость. Скажем,

раньше причиной малярии считался плохой воздух. Теперь мы считаем, что ее переносит малярийный комар. И это имеет значение, потому что, если я захочу пойти на болота, в первом случае мне понадобится дыхательная маска, а во втором — противомоскитная сетка. Мои действия зависят от того, в какую теорию я верю. Переход от одной гипотезы к другой происходит методом проб и ошибок.

Таким способом мы познаем причинно-следственную связь. Но нам нужны способности и шаблон для хранения полученной информации, чтобы ее можно было использовать, тестировать и изменять. Соответственно, мы должны научиться программировать компьютеры таким образом, чтобы они получали этот шаблон и могли им управлять.

М. Ф.: Компания DeepMind использует обучение с подкреплением, которое происходит методом проб и ошибок. Можно ли таким способом обнаружить причинно-следственные связи?

Дж. П.: Обучение с подкреплением позволяет научить только уже известным вещам. Смоделировать гипотетическую ситуацию машина не может.

М. Ф.: Способность прослеживать причинно-следственные связи важна для достижения сильного ИИ?

Дж. П.: Я не сомневаюсь, что это необходимое условие, но не знаю, является ли оно достаточным. Вряд ли причинное осмысление поможет решить все проблемы сильного ИИ. Например, оно не решает проблему распознавания объектов и понимания языка. По сути, задача отслеживания причинно-следственных связей уже решена, что сможет помочь нам в преодолении препятствий, мешающих решить другие задачи.

М. Ф.: Вы считаете, что сильный ИИ возможен?

Дж. П.: Я в этом не сомневаюсь, потому что не вижу теоретических препятствий для его создания.

М. Ф.: Вы упоминали, что разговоры о сильном ИИ велись еще в 1961 г., когда вы работали в RCA. Как вы оцениваете прогресс в этой области? Разочарованы?

Дж. П.: Все идет отлично. Сейчас прогресс замедлен из-за концентрации на глубоком обучении и связанных с ним непрозрачных структурах. Нужно освободиться от философии, ориентированной на данные. В целом на развитие ИИ повлияют люди, а эта сфера привлекает самых умных.

М. Ф.: Но последние успехи в ИИ связаны с глубоким обучением, которое вы критикуете, сравнивая с подгонкой функций к данным и с «черным ящиком», который просто генерирует ответы.

Дж. П.: Да, по сути, сейчас в этой сфере собирают те плоды, которые висят совсем низко.

М. Ф.: Но даже эти результаты весьма впечатляют.

Дж. П.: Это происходит просто потому, что никто не предвидел такого количества низко висящих фруктов.

М. Ф.: Возрастет ли в будущем важность нейронных сетей?

Дж. П.: При правильном использовании нейронные сети и обучение с подкреплением могут стать важными компонентами причинного моделирования.

М. Ф.: То есть, возможно, появится гибридная система, включающая в себя не только нейронные сети, но и идеи из других областей ИИ?

Дж. П.: Гибридные системы создаются уже сейчас, несмотря на малое количество данных. Другое дело, что для получения причин-

но-следственных связей экстраполировать или интерполировать разреженные данные можно только до определенного предела. Даже бесконечные данные не позволяют определить, в чем разница между «А вызывает Б» и «Б вызывает А».

М. Ф.: Если когда-нибудь мы создадим сильный ИИ, будет ли он обладать сознанием и иметь какой-то опыт, подобный человеческому?

Дж. П.: Разумеется, каждая машина имеет своего рода внутренний опыт. Она должна иметь некое представление о своем программном обеспечении, но не может полностью быть в курсе своей архитектуры, так как это противоречит доказанной Тьюрингом проблеме остановки. Но машине целесообразно иметь представление о своих важных связях и модулях. Она должна иметь код своих способностей, убеждений, целей и желаний. И это вполне реально. В каком-то смысле машина уже обладает внутренней сущностью, так как имеет представление об окружающей среде, способна реагировать на происходящее и отвечать на гипотетические вопросы.

М. Ф.: А что вы думаете об эмоциональном опыте? Может ли машина чувствовать себя счастливой или по какой-то причине страдать?

Дж. П.: Это напоминает мне книгу Марвина Минского *The Emotion Machine* («Эмоциональная машина»), в которой он рассказывает, как просто запрограммировать эмоции. В организме присутствуют химические вещества, которые время от времени мешают интеллектуальной деятельности и переопределяют ее. В итоге эмоции можно сравнить с химической машиной, расставляющей приоритеты.

М. Ф.: О чем имеет смысл беспокоиться по мере развития ИИ?

Дж. П.: Мы должны понимать, что именно мы строим. Фактически мы выводим новый вид интеллектуальных животных. Сначала они

будут ручными, как куры или собаки, но в конечном итоге обретут свободу воли, и вот тут следует быть очень осторожными. Но так, чтобы не принести ущерб науке и научному любопытству. Это сложный вопрос, а я бы не хотел вдаваться в тонкости регулирования исследований в сфере ИИ. Это будет если не животное, то пригодный к эксплуатации человек без юридических прав и претензий на заработную плату. Мы вместе работаем над созданием действительно интеллектуальных, гибких ИИ-систем. Осталось встроить перенос обучения.

> *Наш путь кажется мне более органичным, ведь мы работаем над действительно гибкими ИИ-системами для решения новых задач.*

Джефф Дин

СТАРШИЙ НАУЧНЫЙ СОТРУДНИК КОМАНДЫ GOOGLE BRAIN

К компании Google Джефф Дин присоединился в 1999 г., в самом начале ее истории, и за прошедшие годы разработал и реализовал большую часть ее рекламных, поисковых, индексирующих и обслуживающих систем, а также различные части распределенной вычислительной инфраструктуры, на которой работает большинство ее продуктов. Докторскую степень в области computer science получил в Университете штата Вашингтон. В 1996 г. с Крэйгом Чемберсом работал над методиками оптимизации программ для объектно-ориентированных языков. Член Национальной инженерной академии, ACM и Американской ассоциации развития наук (AAAS).

Мартин Форд: Какие исследования в сфере ИИ ведутся в корпорации Google?

Джефф Дин: Наша цель — прогресс в области машинного обучения. Мы пытаемся создавать более интеллектуальные системы, разрабатываем новые алгоритмы, методы, программную и аппаратную инфраструктуру. Google Brain работает над одним из нескольких ИИ-проектов Google и занимается поиском новых областей применения машинного обучения в сотрудничестве с командами Google Translate, Gmail и др. Я лично обеспечиваю энергичную работу по существующим проектам и ищу новые направления.

М. Ф.: В компании DeepMind основное внимание уделяется разработкам сильного ИИ. Означает ли это, что остальные исследования ИИ в Google ориентированы на более узкие и практические приложения?

Дж. Д.: Да, у DeepMind есть структурированный план и вера в себя, но, конечно, о сильном ИИ думают и остальные группы в Google. Наш путь кажется мне более органичным, ведь мы работаем над действительно гибкими ИИ-системами для решения новых задач.

М. Ф.: Как возник ваш интерес к ИИ? И как складывалась ваша карьера?

Дж. Д.: Мне было 9 лет, когда мой отец собрал первый компьютер. К старшим классам школы я уже научился писать программы, а потом получил степень бакалавра в области computer science и экономики в Университете Миннесоты. Темой моей дипломной работы стало параллельное обучение нейронных сетей. Это был 1990 г., когда этой темой интересовались все. Мне очень нравился предоставляемый этими сетями уровень абстракции. Я получал настоящее удовольствие от работы.

Думаю, это чувство испытывали тогда многие, но у нас не было нужных вычислительных мощностей. Мне казалось, что ускорь 64-битные процессорные машины раз в 60, и мы сможем сделать великие вещи. Практика показала, что требовалась в миллион раз большая скорость. Впрочем, теперь она у нас есть.

Затем в течение года я разрабатывал программное обеспечение для статистического моделирования и прогнозирования пандемии ВИЧ/СПИДа во Всемирной организации здравоохранения. После этого в Вашингтоне получил докторскую степень в области computer science. Там я в основном занимался методами оптимизации программ, написанных на объектно-ориентированных языках, в исследовательской лаборатории компании DEC в Пало-Альто.

В конце концов я присоединился к стартапу Google. Тогда в нем работало всего 25 человек. Начал я с разработки нашей первой рекламной системы. Затем был многолетний труд над поисковыми системами и такими функциями, как сканирование, обработка запросов, индексация, ранжирование и т. п., и позднее — инфраструктурное программное обеспечение в проектах MapReduce, Bigtable и Spanner.

В 2011 г. я стал работать над системами, ориентированными на машинное обучение, потому что меня заинтересовало, как можно использовать огромные объемы вычислений, которые применялись для тренировки очень больших и мощных нейронных сетей.

М. Ф.: Расскажите историю проекта Google Brain. Какую роль он играет в Google?

Дж. Д.: Однажды мы разговорились на кухне с Эндрю Ыном, который в то время был консультантом компании Google X, и я узнал, что его ученики в Стэнфорде уже давно интересуются возмож-

ностями нейронных сетей и даже начали решать с их помощью различные проблемы. Меня это заинтересовало, так как с нейронными сетями я работал 20 лет назад, когда писал дипломную работу. В результате у нас родился перспективный план: воспользоваться как можно большим количеством вычислений для обучения нейронных сетей.

Мы решили две проблемы. Во-первых, обучение без учителя на примерах изображений. Мы взяли 10 млн случайных кадров из различных видеороликов YouTube и попытались обучить очень большую сеть. Может быть, вы видели знаменитую визуализацию кошки?

М. Ф.: Да, она наделала много шума.

Дж. Д.: Мы поняли, что при обучении масштабируемых моделей на больших объемах данных происходит что-то интересное.

М. Ф.: Как я понимаю, самое интересное было то, что нейросеть сама извлекла концепцию кота из неструктурированных и немаркированных данных?

Дж. Д.: Да. Мы предоставили сети необработанные изображения и использовали алгоритм обучения без учителя, который позволил восстановить их из сжатой картинки, обнаруживая шаблоны, срабатывающие, когда в центре кадра оказывалась кошка. На YouTube очень много кошек. Во-вторых, вместе с командой по распознаванию речи мы попробовали применить глубокое обучение и нейронные сети для создания акустической модели перехода от необработанных звуковых сигналов к частям слова, например, «buh», «fuh» или «ss». Нейронные сети превзошли своих предшественников по точности распознавания речи.

Затем с другими командами Google мы занялись восприятием, параллельно применяя наши подходы к новым задачам. Наши

системы представляли собой большие вычисления на нескольких компьютерах, которые не требовали привлечения программиста. Достаточно было указать: «Я хочу обучить вот эту большую модель, пожалуйста, используйте сто компьютеров». Это было первое поколение программного обеспечения для подобных задач.

Затем появилось второе поколение, то есть библиотека TensorFlow, которую мы решили сделать системой с открытым исходным кодом. Проектируя ее, мы преследовали три цели. Во-первых, гибкость, позволяющую быстро проверять различные идеи в области машинного обучения. Во-вторых, возможность масштабирования. В-третьих, мы хотели перейти от исследований к системе обслуживания производства. С 2015 г. ее применяют многие компании, научные учреждения и обычные люди.

М. Ф.: Собираетесь ли вы добавить TensorFlow на ваш облачный сервер, чтобы дать своим клиентам доступ к машинному обучению?

Дж. Д.: Да, собираемся, но мы хотим сперва сделать облако лучшим местом для запуска и использования библиотеки где угодно: на ноутбуке, стационарном компьютере, Raspberry Pi и Android.

М. Ф.: Облачная служба Google Cloud будет предоставлять тензорные процессоры и специализированное оборудование для оптимизации?

Дж. Д.: Да, одновременно с разработкой TensorFlow мы создавали процессоры, которым доступны операции из линейной алгебры, составляющие ядро всех новейших приложений глубокого обучения. Их отличает высокая скорость работы, низкие энергозатраты, увеличение пропускной способности Google Translate, Google Search и систем распознавания речи. Кроме того, для облачных

клиентов будут доступны тензорные процессоры второго поколения (TPU). Пользователь сможет создать виртуальную машину, имеющую доступ к аппаратной части TPU, и запускать на ней приложения из TensorFlow.

М. Ф.: Означает ли интеграция этих технологий в облако, что мы приближаемся к общедоступному машинному обучению?

Дж. Д.: У нас есть множество облачных продуктов для опытных пользователей и новичков.

Достаточно отправить нам изображение или аудиоклип, и мы расскажем, что там содержится. Например, «это изображение кота», или «люди на картинке выглядят счастливыми», или «мы извлекли из изображения следующие слова», «мы думаем, что в этом аудиоклипе говорится следующее».

Тем, кто хочет получить индивидуальное решение конкретной проблемы, мы предлагаем набор продуктов с различными стадиями AutoML. Он может из набора изображений сотни видов деталей сборочной линии идентифицировать конкретную деталь.

Мне это кажется важным. По моим прикидкам, в мире существует 10–20 тысяч организаций, в которых есть специалисты по машинному обучению. И миллионов десять компаний, у которых есть пригодные для машинного обучения данные и проблемы, которые можно решить с помощью этой технологии.

Наша цель — сделать так, чтобы применение методов машинного обучения стало таким же простым, как отправка запроса к базе данных. Тогда машинное обучение можно будет использовать в небольших городах для установки таймеров светофоров.

М. Ф.: То есть вы стремитесь к демократизации ИИ. А что, на ваш взгляд, мешает созданию сильного ИИ?

Дж. Д.: При обучении с учителем после постановки задачи начинается сбор данных. В итоге появляется модель, которая хорошо справляется с конкретной задачей, но не умеет ничего другого.

Для получения универсальной интеллектуальной системы нужна модель, способная выполнять сотни тысяч задач. Которая, когда перед ней будет поставлена 100 001-я задача, сможет самостоятельно использовать накопленный опыт для разработки новых, эффективных методов ее решения. Чтобы постепенно для решения новых задач требовалось все меньше и меньше данных и наблюдений.

Я думаю, это можно реализовать путем экспериментов. Возможно, системы могут учиться на демонстрируемых примерах. По сути, это тоже обучение на маркированных данных, просто роботам показывается, скажем, как человек наливает жидкость в стакан, а они повторяют это действие. По идее в таком случае для обучения будет достаточно небольшого количества примеров.

Но для создания таких систем требуется масса вычислительных ресурсов. Ведь, чтобы попробовать разные подходы к решению задач, эксперименты должны проводиться очень быстро. Это одна из причин для инвестиций в крупномасштабные аппаратные ускорители машинного обучения, такие как TPU.

М. Ф.: А какие опасности, с вашей точки зрения, несет с собой ИИ? О чем нам имеет смысл беспокоиться?

Дж. Д.: Правительствам следует готовиться к крупным изменениям на рынках труда. Уже сейчас компьютеры могут автоматизировать многие вещи, которые были недоступны для автоматизации лет пять назад. И постепенно этот процесс затронет множество различных профессий и рабочих мест.

В 2016 г. я присутствовал на заседании Бюро по определению научно-технической политики Белого дома. Там примерно 20 спе-

циалистов по машинному обучению и около 20 экономистов обсуждали будущее рынков труда. Потому что это задача правительства — выяснить, что можно предложить людям, чьи рабочие места меняются или исчезают, какие новые навыки или варианты переобучения доступны.

М. Ф.: Потребуется ли нам в один прекрасный день такая вещь, как универсальный базовый доход?

Дж. Д.: Я не знаю. Трудно предсказать, как будут развиваться события, потому что сколько бы связанных с развитием технологий изменений мы ни переживали в прошлом, каждый раз это происходило по-новому. И промышленная революция, и сельскохозяйственная революция вызывали дисбаланс в обществе.

Людям уже сейчас важно проявлять гибкость и изучать новые вещи. Полвека назад выпускник вуза мог начать карьеру и долгое время ее строить, сегодня же обычна ситуация, когда человек, поработав несколько лет, приобретает новые навыки и уходит в другую сферу.

Что касается других рисков, меня, в отличие от Ника Бострома, не очень беспокоит возможность появления суперинтеллекта. Я уверен, что, как ученые и исследователи, мы сможем создать системы машинного обучения таким образом, чтобы они интегрировались в общество и начали приносить пользу всем. К сожалению, прекрасные вещи сопровождаются многочисленными нюансами, о которых не стоит забывать.

М. Ф.: Но ведь разработкой сильного ИИ занимается небольшая группа людей, которая не обязана быть в курсе всех сопутствующих проблем. Должны ли регулироваться такие исследования и применение ИИ?

Дж. Д.: Возможно. Но хотелось бы, чтобы этим занимались люди, обладающие нужным опытом. Мне кажется, что регулирование

порой происходит с задержкой, так как государственные структуры пытаются угнаться за стремительно меняющейся ситуацией. В таких вопросах нужна не реакция под влиянием момента, а информированный диалог с людьми на местах.

Разрабатывая сильный ИИ, важно помнить об этических аспектах. Именно поэтому Google выпустила документ, в котором четко и недвусмысленно перечислены принципы, которыми руководствуется в исследованиях[1]. Мы думаем не только о техническом развитии, но и о том, какие проблемы мы хотим решать и какими способами.

[1] https://www.blog.google/technology/ai/ai-principles/.

> *Если мы откажемся от технологического прогресса, его осуществит кто-то другой, возможно, имеющий не такие мирные намерения.*

Дафна Коллер

ГЕНЕРАЛЬНЫЙ ДИРЕКТОР И ОСНОВАТЕЛЬ КОМПАНИИ INSITRO, ВНЕШТАТНЫЙ ПРЕПОДАВАТЕЛЬ COMPUTER SCIENCE В СТЭНФОРДЕ

Дафна Коллер — бывший профессор computer science в Стэнфорде. Вместе с Эндрю Ыном основала образовательную онлайн-компанию Coursera и стала ее президентом. Была директором по информационным технологиям компании Calico. Внесла значительный вклад в область байесовского машинного обучения и представления знаний. Степени бакалавра и магистра получила в Еврейском университете в Израиле, а степень доктора наук в области computer science — в Стэнфорде. Имеет множество наград, в том числе стипендию Макартура. Член AAAI и Национальной инженерной академии. В 2013 г. попала в список 100 самых влиятельных людей мира по версии журнала Time.

Мартин Форд: Расскажите о вашем стартапе insitro.

Дафна Коллер: Мы в insitro надеемся, что большие данные и машинное обучение вкупе с новейшими медико-биологическими разработками помогут ускорить, удешевить и сделать более эффективным процесс разработки лекарственных средств.

Я начала заниматься машинным обучением в сфере здравоохранения 17 лет назад, когда большим считался набор данных из нескольких десятков выборок. Сейчас британский UK Biobank хранит миллионы биообразцов. При этом технологии позволяют создавать модели биологических систем и экспериментировать с ними с беспрецедентной точностью и пропускной способностью.

Без наших разработок показатели успешности клинических испытаний крайне низки: стоимость создания каждого нового лекарства (с учетом неудачных попыток) оценивается более чем в 2,5 млрд долларов, а доходность от инвестиций в эту сферу постоянно снижается и согласно некоторым расчетам станет нулевой к 2020 г. Происходящее объясняют, в частности, тем, что уже найдены лекарства, которые эффективно работают для большинства, и впереди более сложная работа.

М. Ф.: То есть компания insitro планирует совмещать экспериментальные работы с машинным обучением. Как это реализовать в рамках одной компании?

Д. К.: Труднее всего заставить ученых и специалистов по данным работать как равные партнеры. Во многих компаниях направление задает одна группа, а все остальные отходят на второй план. В insitro мы пытаемся сформировать культуру, в которой ученые, инженеры и специалисты по данным работают в тесном и равноправном сотрудничестве.

М. Ф.: Насколько важно машинное обучение в сфере здравоохранения?

Д. К.: Машинное обучение имеет смысл применять в сферах, где можно накопить большие объемы данных. При этом нужно, чтобы специалисты в проблемной области одновременно разбирались в машинном обучении и понимали, как с помощью этой технологии решать имеющиеся задачи.

Сегодня без проблем можно получить большой объем данных из таких ресурсов, как UK Biobank или All of Us. Кроме того, появились удивительные технологии, например, CRISPR, синтез ДНК, секвенирование нового поколения и другие вещи, позволяющие создавать большие наборы данных на молекулярном уровне.

Уже можно начинать деконволюцию одной из самых сложных систем: биологии людей и других организмов. Для науки это невероятная возможность.

М. Ф.: Как возник ваш интерес к ИИ?

Д. К.: В аспирантуре в Стэнфорде я занималась вероятностным моделированием. Сейчас бы сказали, что я занимаюсь ИИ, но в сфере ИИ в то время в основном фокусировались на логических рассуждениях.

Потом я стала работать в Беркли и задумалась о связи моих разработок с реальными проблемами. В 1995 г. я вернулась в Стэнфорд как преподаватель и занялась вещами, связанными со статистическим моделированием и машинным обучением, подбирая к ним прикладные задачи.

Какое-то время я работала в области компьютерного зрения и робототехники, а с 2000 г. перешла в сферу здравоохранения. Меня всегда интересовало обучение с применением новейших технологий, в результате чего в 2011 г. появились три массовых открытых онлайн-курса, которые вирусным образом приобрели невероятную популярность. Это дало толчок к созданию проекта Coursera.

М. Ф.: Прежде чем мы начнем разговор об этом проекте, я хочу узнать о ваших исследованиях. Вы занимались байесовскими сетями и интеграцией вероятностного подхода в машинное обучение. Эти вещи можно использовать с нейронными сетями глубокого обучения?

Д. К.: На этот вопрос нет однозначного ответа. Вероятностные модели лежат между моделями, которые кодируют предметные знания интерпретируемым образом (имеющим смысл для людей), и моделями, которые передают статистические свойства данных. Глубокое обучение направлено на точность прогнозирования, зачастую в ущерб интерпретируемости, столь важной в медицинских приложениях. Хотя, конечно, способность обходиться без предварительно накопленных знаний, извлекая всю необходимую информацию из данных, имеет много преимуществ. Было бы здорово соединить все это.

М. Ф.: Расскажите о проекте Coursera.

Д. К.: Мы изо всех сил пытались выяснить, в каком направлении лучше всего двигаться. Стоит ли продолжить деятельность на базе Стэнфорда? Или создать некоммерческую организацию? Или открыть компанию? В итоге нам показалось, что максимальный результат даст создание компании. И в январе 2012 г. появилась Coursera.

М. Ф.: Я помню весь этот шум в СМИ о том, что теперь для обучения в Стэнфорде достаточно мобильного телефона. Мне кажется, ажиотаж в изрядной степени был вызван людьми, которые, уже имея высшее образование, посещали сайт Coursera с целью повышения квалификации. Ведь вопреки многочисленным предсказаниям студенты продолжают обучаться стандартным способом. Или со временем ситуация может поменяться?

Д. К.: Первым делом подчеркну, что мы никогда не позиционировали себя как замену полноценного высшего образования.

Разумеется, разговоры на эту тему периодически возникали, но лично я не считаю эту идею хорошей. В 2012 г. утверждалось, что массовые онлайн-курсы оставят университеты без работы. Спустя год начались разговоры о том, что раз университеты все еще функционируют, идея массовых открытых онлайн-курсов провалилась. Оба утверждения представляют собой нелепые крайности, типичные для шумихи.

На мой взгляд, мы многое сделали для людей, лишенных доступа к высшему образованию. Около 25 % обучающихся на сайте Coursera не имеют научных степеней, а около 40 % проживают в развивающихся странах. Учащиеся, которые говорят, что их жизнь сильно изменилась благодаря нашему ресурсу, по большей части имели низкий социально-экономический статус. Разумеется, чтобы воспользоваться онлайн-образованием, нужен доступ в интернет. Кроме того, человек должен знать, что такая возможность существует. Надеюсь, со временем мы сможем повысить степень осведомленности.

М. Ф.: Есть мнение, что мы склонны переоценивать происходящее сейчас и недооценивать вещи, имеющие смысл в долгосрочной перспективе. Мне кажется, сейчас именно такой случай.

Д. К.: Я с вами согласна. Люди думали, что за пару лет мы трансформируем высшее образование. При этом университеты существуют около 500 лет и развиваются медленно. Но мне кажется, что даже за пять лет нашей работы многое изменилось.

Например, университеты стали предлагать онлайн-курсы зачастую по значительно более низкой цене, чем очное обучение. Когда Coursera только появилась, невозможно было представить, что у университетов появятся онлайн-программы. А сейчас цифровое обучение стало обычной вещью.

М. Ф.: Стэнфорду в ближайшее десятилетие вряд ли что-то угрожает, но обучение в других, менее престижных вузах все еще стоит очень дорого. Тогда люди начнут выбирать Coursera?

Д. К.: Думаю, что преобразованию в первую очередь подвергнется сфера последипломного образования, в частности магистратура. У очной формы обучения есть социальный компонент: общение, перемена места жительства. В аспирантуре же обычно учатся уже взрослые люди, имеющую семью и работу. В дальнейшем задумываться о более эффективном использовании времени и денег начнут студенты небольших колледжей, совмещающие учебу с работой.

М. Ф.: В рамках курсов находится огромное количество данных, пригодных для машинного обучения и ИИ. Можно ли интегрировать эти технологии в ваши курсы, сделав их более динамичными, более персонализированными и т. п.?

Д. К.: Думаю, да. В начале проекта мы пользовались стандартными методиками обучения. Интерактивность курсов обеспечивалась упражнениями. По мере сбора данных обучение может стать более персонализированным. Возможно, появится что-то вроде личного репетитора.

М. Ф.: Что вы думаете о перспективах глубокого обучения?

Д. К.: Глубокое обучение — это не волшебная палочка, но и отказываться от него не стоит. Оно позволило сделать важный шаг вперед, но вряд ли поможет достичь ИИ человеческого уровня. Нужен, по крайней мере, еще один крупный прорыв.

Сейчас сети оптимизируются под конкретные задачи, но под другие задачи их приходится переобучать или даже менять архитектуру. Фактически сейчас мы сосредоточены на узких проблемах, решения которых не пересекаются друг с другом.

Кроме того, для обучения таких моделей требуется огромный набор данных, в то время как людям хватает и небольшого объема информации. Я подозреваю, что в нашем мозге есть некая

структура, отвечающая за выполнение всех задач с той скоростью, которая для робота или агента недоступна.

М. Ф.: Какие еще препятствия стоят на пути к сильному ИИ? Как добавить роботам воображение и способность придумывать новые идеи?

Д. К.: Алгоритм GAN умеет создавать изображения, отличные от ранее увиденных. Но это сплав изображений из обучающего набора.

Трудно проверить наличие некоторых свойств. Люди могут притворяться, имитировать эмоциональную связь с другими и научить этому компьютер. Тест Тьюринга определяет, что поведение другого существа совпадает с поведением, которое мы считаем сознательным, но границы сознания не так четко определены.

М. Ф.: Хороший вопрос, подразумевает ли сильный ИИ наличие сознания или же это сверхинтеллектуальный зомби? Возможна ли невероятно умная машина, совершенно лишенная внутренних переживаний?

Д. К.: Согласно гипотезе, ставшей основанием для теста Тьюринга, сознание непостижимо. Например, я не знаю, обладаете ли вы сознанием, но принимаю этот факт на веру, ведь вы похожи на меня, а себя я ощущаю существом, обладающим сознанием.

Тьюринг утверждал, что по поведению сущности, которая достигла определенного уровня производительности, невозможно определить, обладает она сознанием или нет. Соответственно, получается, что вопрос о сознании у машин навсегда останется открытым.

М. Ф.: Расскажите о последних разработках в сфере ИИ.

Д. К.: Комбинация доступных в настоящее время объемов данных с глубоким обучением позволяет машине самостоятельно выбирать шаблоны. Поскольку архитектура модели зависит от предметной области, до сих пор не появились компьютеры, самостоятельно проектирующие сети глубокого обучения. Я думаю, что в этом деле человек незаменим. Хотя предпосылки к изменению ситуации тоже есть, благодаря сквозному обучению в сочетании с неограниченными тренировочными данными, которые уже использованы в AlphaGo и AlphaZero.

М. Ф.: Как вы считаете, сколько времени потребуется на создание сильного ИИ и по каким признакам можно понять, что мы приближаемся к этому моменту?

Д. К.: Для сильного ИИ требуется ряд крупных технологических прорывов. Предсказать, когда они произойдут, невозможно.

М. Ф.: Но если эти прорывы совершатся, мы быстро придем к сильному ИИ?

Д. К.: Даже в этом случае для воплощения сильного ИИ в жизнь потребуется много технических расчетов и работы. Возможно, ключевое открытие уже сделано, мы об этом просто не знаем. И нужны десятилетия работы, чтобы спроектировать все в окончательном виде.

М. Ф.: А ведь нельзя забывать об опасностях, которые несет с собой ИИ, например, в сфере экономики. Как вы думаете, ждет ли нас обвал рынка труда?

Д. К.: В ситуации, когда множество человеческих обязанностей полностью или частично возьмут на себя машины, это неизбежно. Во многих случаях этому процессу мешают социальные препятствия, но значительный рост производительности приведет к тому, что рано или поздно все равно начнется цикл разрушительных инноваций.

Помощники юристов и кассиры в супермаркетах уже понемногу заменяются машинами, скоро очередь дойдет до тех, кто выкладывает товар на прилавки. Лет через пять (максимум — десять) эти обязанности возьмут на себя роботы или интеллектуальные агенты. Вопрос в том, в какой степени мы сможем создать для людей новые значимые рабочие места.

М. Ф.: Как скоро, на ваш взгляд, начнут массово использоваться беспилотные автомобили?

Д. К.: Мне кажется, что в этой сфере произойдет постепенный переход. Сначала за автомобилями будут наблюдать водители-люди. Это неизбежный промежуточный шаг на пути к полной автономии.

Сидящий в офисе водитель будет контролировать три или четыре автомобиля одновременно. Оказавшись в ситуации, требующей нестандартных действий, машина может позвать его на помощь. Беспилотные автомобили в таком формате могут начать предоставляться как услуга лет через пять. Для полной же автономии нужна скорее социальная, а не техническая эволюция. Предсказать сроки ее наступления куда сложнее.

М. Ф.: Но даже в описанном вами варианте множество водителей потеряет работу. Можно ли решить эту проблему введением универсального базового дохода?

Д. К.: Пока об этом рано говорить. Если вспомнить историю, то и раньше, во время сельскохозяйственной, а потом и промышленной революции делались прогнозы по поводу массовых увольнений и огромного числа людей, оставшихся без работы. Мир изменился, и люди нашли другую работу.

Нужно думать об отношении к образованию. Инвестиции в переобучение явно недостаточны для того, чтобы люди могли найти себе место в новой реальности. Важно понять, какие навыки нуж-

ны людям для успешного продвижения вперед. И только если это не даст возможности сохранить занятость большинства населения, думать о всеобщем базовом доходе.

М. Ф.: С ИИ связывают и другие риски. Их можно разделить на две категории: краткосрочные, такие как проблемы конфиденциальности, безопасности и использования дронов, и долгосрочные, такие как собственные цели сильного ИИ.

Д. К.: Я бы сказала, что все вещи, перечисленные в первой группе, существуют и без привязки к ИИ.

М. Ф.: Но разве риски не растут по мере расширения технологий? Представьте, что грузовики автономно развозят по магазинам продукты и вдруг кто-то взламывает систему и останавливает их движение.

Д. К.: Я согласна, что мы все больше полагаемся на электронные решения, у которых в силу их взаимосвязи возрастает вероятность появления единой точки отказа. Сейчас, чтобы парализовать доставку товаров в магазины, нужно вывести из строя всех водителей.

М. Ф.: Вас беспокоит возможность злонамеренного использования технологии?

Д. К.: Конечно, любая технология может попасть в недобрые руки. Человечество постоянно совершенствует способы убийства себе подобных. Оружие массового поражения появилось довольно давно. Сейчас говорят, например, о биологическом оружии, о возможности создавать новые вирусы, действующие только на представителей определенных групп. И я бы не сказала, что оснащенные ИИ беспилотники-убийцы более опасны, чем люди, которые синтезируют новые штаммы оспы.

М. Ф.: Давайте рассмотрим риски, связанные с сильным ИИ: проблему выравнивания, появление собственной цели и ее

реализацию неожиданным для нас или даже вредоносным способом.

Д. К.: Я считаю, что думать обо всех этих вещах несколько преждевременно. Когда в процессе проектирования этой системы мы выясним, из каких ключевых компонентов она состоит, тогда придет время задуматься, как получить самый лучший результат.

М. Ф.: Но этой проблемой уже занимается ряд организаций, например компания OpenAI. С вашей точки зрения, это преждевременные инвестиции?

Д. К.: Компания OpenAI занимается не только этим вопросом. Например, они создают инструменты ИИ с открытым исходным кодом для демократизации доступа к ценной технологии. И мне кажется, что это здорово. На конференции по машинному обучению NIPS 2017 был интересный доклад о том, как машинное обучение усиливает неявные искажения в тренировочных данных до такой степени, что начинают проявляться худшие варианты поведения (например, расизм или сексизм). Вот о чем важно думать уже сейчас, потому что это реальная опасность и нужно искать способ ее устранения.

М. Ф.: Нужны ли сфере ИИ регуляционные меры со стороны правительства?

Д. К.: Скажем так, я думаю, что уровень понимания этой технологии членами правительства в лучшем случае ограничен. Не очень хорошая идея регулировать вещи, в которых не разбираешься.

Кроме того, такая технология, как ИИ, проста в использовании и уже доступна правительствам других стран, которые не обязательно связаны теми же этическими нормами, что и мы. Так что регулирование этой технологии, на мой взгляд, не имеет особого смысла.

М. Ф.: Да, сейчас много говорят о преимуществах Китая, в котором благодаря количеству населения собираются огромные объемы данных, кроме того, там не так сильно беспокоятся о конфиденциальности. Как вы считаете, есть ли у нас риск отстать?

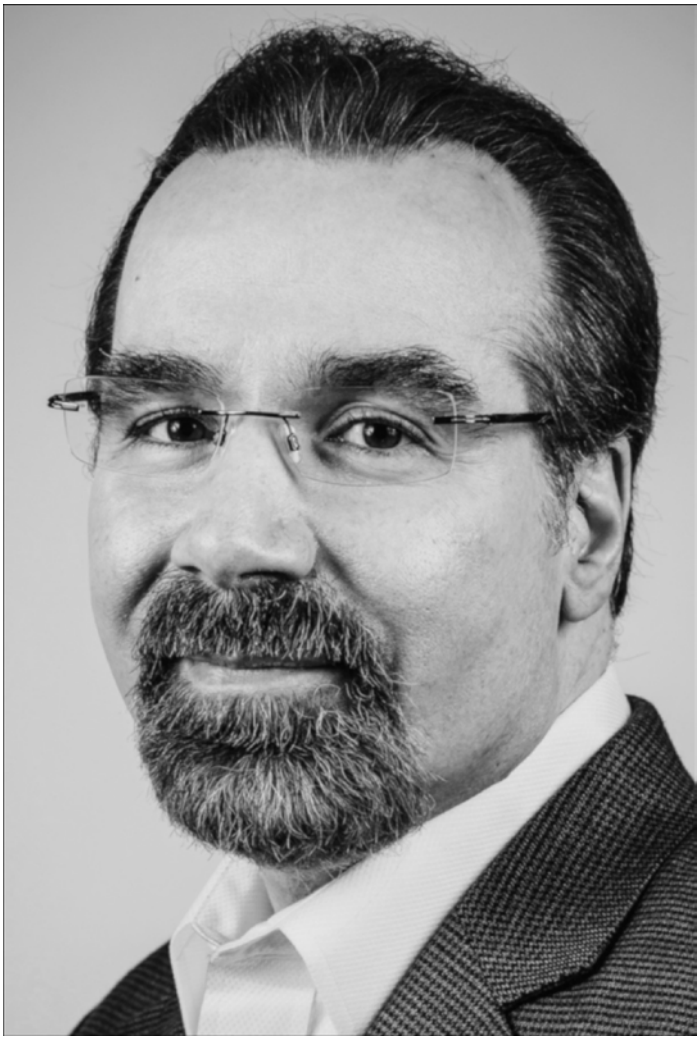
Д. К.: Такой риск присутствует. И если говорить о вмешательстве со стороны правительства, то было бы полезно, если бы на государственном уровне поспособствовали технологическому прогрессу, чтобы поддержать нашу конкурентоспособность. Это инвестиции в науку, образование и конфиденциальность при доступе к данным.

Например, многое можно сделать в сфере здравоохранения. Большинство пациентов с удовольствием предоставит данные для исследований, направленных на борьбу с их заболеваниями. Люди понимают, что даже если они не получат помощи, в будущем это может помочь другим, и хотят в этом участвовать. Но сейчас правовые и технологические препоны, которые нужно преодолеть для распространения медицинских данных, настолько суровы, что зачастую с ними даже не пытаются бороться. Это замедляет поиск способов лечения и другие исследования.

Именно в этой сфере может иметь значение как изменение политики на государственном уровне, так и изменение социальных норм. Например, в одних странах люди подписывают согласие на донорство органов после смерти, а в других — отказ. В обоих случаях есть государственный контроль, но там, где граждане выражают согласие, количество доноров выше. Потому что отказ создает ощущение принуждения. Если использовать подобную систему для обмена данными, новые исследования будут происходить быстрее.

М. Ф.: Вы верите в то, что выгоды от ИИ, машинного обучения и остальных технологий перевешивают сопутствующие риски?

Д. К.: Да. Чтобы уменьшить риски, нужно думать о мерах предосторожности и изменении социальных норм. Если мы откажемся от технологического прогресса, его осуществит кто-то другой, возможно, имеющий не такие мирные намерения. Технологии должны развиваться, а нам нужно думать о механизмах, позволяющих направить их в позитивную сторону.



> *В отличие от других, я не считаю, что для достижения сильного ИИ нужно ждать какого-то огромного прорыва. Думаю, мы уже знаем, как действовать, просто нужно доказать эффективность этого пути.*

Дэвид Ферруччи

ОСНОВАТЕЛЬ И ГЕНЕРАЛЬНЫЙ ДИРЕКТОР КОМПАНИИ ELEMENTAL COGNITION LLC, СОТРУДНИК КОМПАНИИ BRIDGEWATER ASSOCIATES

Дэвид Ферруччи был ведущим специалистом IBM по системе Watson, которая в 2011 г. одержала победу в телевикторине «Своя игра». Окончил Манхэттенский колледж со степенью бакалавра биологии и Политехнический институт Ренсселера со степенью доктора computer science. Специализировался на представлении знаний и аргументации. Имеет более 50 патентов, опубликовал многочисленные статьи на тему ИИ, автоматических рассуждений, НЛП, архитектур интеллектуальных систем, автоматического создания историй и автоматического ответа на вопросы.

Получил звание Почетного сотрудника IBM и множество наград за создание стандарта UIMA и суперкомпьютера Watson. В их числе премия Чикагской товарной биржи за инновации и премия Фейгенбаума от AAAI.

Мартин Форд: Как возник ваш интерес к компьютерам и ИИ?

Дэвид Ферруччи: Мои родители хотели, чтобы я стал врачом. Папе очень не нравилось, что я во время школьных каникул болтаюсь без дела. И после окончания предпоследнего класса он через газету нашел для меня уроки математики в местном колледже. Но оказалось, что на самом деле там учат программировать компьютеры DEC на языке BASIC. Меня восхитило, что можно придумать алгоритм, дать машине инструкции, а дальше она сделает все сама. Это показалось мне спасением! Если заставить машину все запоминать и думать вместо меня, я смогу без усилий стать врачом.

Термина «ИИ» я в то время не знал, но меня интересовало представление интеллекта с математической, алгоритмической и философской точки зрения. Я верил в возможность смоделировать интеллект человека на компьютере.

М. Ф.: И в колледже вы начали изучать computer science?

Д. Ф.: Нет, я не представлял себя в сфере computer science или ИИ. В колледже я специализировался в биологии. Однажды попросил дедушку купить мне компьютер Apple II и начал программировать все, что приходило в голову: от графического программного обеспечения для лабораторных работ до экологического симулятора и аналого-цифрового интерфейса для лабораторного оборудования. Готовых программ тогда не было. В последний год обучения я так много занимался computer science, что это стало моей второй специальностью. Я получил высокие оценки по биологии и даже собирался поступать в медицинский институт, но внезапно понял, что это не мое.

Вместо этого я поступил в магистратуру по специальности «computer science и ИИ» в Политехнический институт Ренсселера в Нью-Йорке. В рамках диссертации я разработал семантическую сетевую систему, которую назвал COSMOS. Точной расшифровки аббревиатуры я не помню. Система генерировала представление

знаний и языка и до некоторой степени могла рассуждать логически.

В 1985 г. на ярмарке технических наук моего вуза я проводил презентацию системы COSMOS. Там меня увидели ребята из Исследовательского центра IBM, которые недавно начали проект по ИИ. Они спросили, не нужна ли мне работа. Я в то время планировал окончить учебу и получить научную степень, но из прочитанной ранее газетной статьи знал, что научные сотрудники IBM могут исследовать что угодно с неограниченными ресурсами. Эта статья описывала работу моей мечты, поэтому я вырезал ее и прикрепил на свою пробковую доску. И мне ее предложили!

В 1985 г. я приступил к работе над ИИ-проектом в подразделении IBM Research, а через пару лет началась «зима ИИ». Мне сказали, что я могу присоединиться к другим проектам, но я хотел работать только над ИИ и ушел из IBM. Папа страшно злился на меня: сначала я не стал врачом, а теперь потерял хорошую работу.

Я вернулся в Политехнический институт Ренсселера, где защитил диссертацию по немонотонной логике и спроектировал медицинскую систему CARE (cardiac and respiratory expert — эксперт по болезням сердца и дыхательных путей). Чтобы оплачивать обучение, я работал над объектно-ориентированной системой проектирования схем по государственному контракту. После получения научной степени я стал искать работу, и в это время сильно заболел мой отец. Он жил в Уэстчестере, где базировалась фирма IBM, поэтому я вернулся в IBM Research.

На тот момент ИИ там совсем не занимались, но 15 лет спустя, благодаря суперкомпьютеру Watson и другим проектам, исследования в этом направлении возобновились. Я никогда не отказывался от своего желания работать над ИИ, постепенно создавая квалифицированную команду. Когда возник интерес к телевикторине «Своя игра», я был единственным в фирме, кто верил в наши силы.

М. Ф.: История этого суперкомпьютера уже очень хорошо документирована, и я бы не хотел на ней останавливаться, поэтому лучше расскажите, что вы думали об ИИ после ухода из IBM.

Д. Ф.: С моей точки зрения, основу для коммуникации, а также развития теорий и идей составляют такие вещи, как восприятие и распознавание объектов, контролирование и выполнение различных задач, а также знание, построение, развитие и понимание концептуальных моделей.

В процессе работы над проектом Watson я узнал, что чисто статистические подходы ограничены в возможности давать ответы на произвольные вопросы. Статистические подходы эффективны для задач восприятия, таких как распознавание образов и голоса, а также для задач управления, например, беспилотными автомобилями и роботами, но в понимании пространства знания ИИ буксуют.

М. Ф.: В 2015 г. вы основали собственную компанию Elemental Cognition. Расскажите о ней.

Д. Ф.: Это исследовательская компания, занимающаяся ИИ, которая пытается добиться реального понимания языка. Это область ИИ, в которой пока ничего особого не сделано. Мы хотим выйти за пределы поверхностной структуры языка и шаблонов, которые встречаются в частоте повтора слов, и понять скрытый смысл. Это позволит строить внутренние логические модели, которыми люди пользуются для рассуждений и общения. Мы хотим создать систему, автономно изучающую окружающий мир, в том числе через диалог с человеком.

Попытки определить, что же такое знание и понимание, — интересная часть нашей работы, ведь люди могут давать разные интерпретации одного и того же текста.

М. Ф.: Да, распознав на изображении кошку, существующие системы не воспринимают ее так же, как люди.

Д. Ф.: Вопрос «что такое кошка?» так же труден, как «что такое понимание?». Вспомните, сколько энергии уходит на то, чтобы достичь взаимопонимания между людьми. Поэтому в науках разработаны формальные языки, которые недвусмысленно передают информацию. Для обычного общения мы используем естественный язык, в котором важны контексты и намерения. Чтобы достичь уверенного понимания, люди задают друг другу вопросы, возвращаются к ранее сказанному, согласовывают свои представления, пока у обеих сторон не появятся похожие модели предмета разговора. И все потому, что сам язык не несет информации, а является только средством синхронизации самостоятельных моделей.

Например, когда моей дочери было семь лет, она прочитала в книге, что электричество — это энергия, которая создается разными способами, в частности с помощью воды, вращающей турбины. После текста задавался простой вопрос: «Как производится электричество?». Дочь еще раз заглянула в текст, сопоставила его с вопросом и сказала, что слово «создано» — синоним слова «произведено», а значит, ответом будет фраза: «Вода вращает турбины». Примерно так работает большинство современных языковых ИИ-приложений. Разница в том, что моя дочь осознавала, что сути она не понимает, а только копирует информацию. Ей было интересно, как все выглядит на самом деле, потому что от своего логического представления она ожидала гораздо большего. Я воспринял это как признак интеллекта.

М. Ф.: Вы говорите сейчас об ИИ человеческого уровня?

Д. Ф.: Когда мы научимся создавать системы, умеющие обучаться автономно, то есть понимать прочитанное и строить на его основе модели, а затем синхронизировать их в процессе общения, можно будет говорить, что мы приблизились к тому, что я назвал бы целостным интеллектом.

Я делю полный ИИ на три части: восприятие, контроль и знание. Многие из процессов, происходящих в сфере глубокого обучения,

позволили значительно продвинуться в первых двух элементах. Третьим элементом мы пытаемся заниматься в компании Elemental Cognition.

М. Ф.: Решить проблему понимания — одна из заветных целей всех, кто занимается ИИ. После этого многое станет возможным. Например, перенос обучения — применение знаний из одной области в другой.

Д. Ф.: Совершенно верно. Среди прочего мы занимаемся тестированием того, как система понимает и объединяет сведения, которые она в процессе чтения извлекла даже из самой простой истории. Может ли она, прочитав рассказ о футболе, разобраться, что происходит во время игры в баскетбол? Как она использует концепции, с которыми познакомилась? Может ли проводить аналогии?

Сложность в том, что люди рассуждают разными способами. Они делают то, что можно сравнить со статистическим машинным обучением. То есть обрабатывают множество данных, формируют обобщенный шаблон и применяют его. Они формируют в голове своего рода линию тренда и с ее помощью интуитивно ищут новые ответы. Могут взглянуть на набор значений и сказать, какое значение может быть следующим. По сути, происходит сопоставление с образцом и экстраполяция. Причем обобщение может оказаться более сложным, чем в случае с линией тренда. Это сравнимо с методами глубокого обучения.

Но когда человек говорит: «Я считаю, что следующим в этом наборе чисел должно быть число 5, потому что...», у него появляется логическая или причинная модель. Это уже совсем другой, гораздо более полезный вид информации, потому что появляется возможность аргументированно возразить: «Вот здесь твои рассуждения ошибочны, потому что…». Подобное невозможно в ответ на заявление: «Это просто моя интуиция, основанная на данных. Доверься мне».

Мы сталкиваемся с дилеммой, противопоставляя эти два вида интеллектуальной обработки данных. Ту, которая связана с построением объяснимой модели, доступной для осмотра, обсуждения и совершенствования, и ту, которая утверждает: «Я считаю так, потому что…». Полезны и необходимы оба вида. Мир под управлением машин, которые не могут объяснить ход своих рассуждений, мне не нравится.

М. Ф.: Но многие считают, что для движения вперед достаточно глубокого обучения, которое соответствует второй из описанных вами моделей. Вы же, как я понимаю, говорите о необходимости других подходов.

Д. Ф.: Я не фанатик какого-то единственного подхода. Глубокое обучение и нейронные сети успешно обнаруживают нелинейные, очень сложные зависимости в больших объемах данных. Поведение сложных систем зачастую описывается функциями многих переменных, которые могут иметь разрывы во многих измерениях.

Системе глубокого обучения можно предоставить огромный объем необработанных данных, и она найдет сложную функцию, но изучать ее в итоге придется вам. Вы можете, конечно, возразить, что любая форма интеллектуальной обработки данных, по сути, сводится к изучению функций. Но если вы не попытаетесь найти функцию, которая реализует сам человеческий интеллект, вы не поймете, почему система дает именно такие ответы.

Можно ли обучить нейронную сеть на данных, присваивающих каждому английскому рассказу его значение? Возможно, да. Но такие данные у нас отсутствуют, и мы не знаем, как изучать их с точки зрения сложности функции. Я не знаю, как при современном состоянии нейронных сетей должна выглядеть методология получения такой информации. Кое-какие идеи у меня, конечно, имеются. И для этого я использую и нейронные сети, и другие методы машинного обучения как часть комплексной архитектуры.

М. Ф.: В документальном фильме «Доверяете ли вы этому компьютеру?» вы сказали, что через три-пять лет появится компьютерная система, которая найдет общий язык с человеком. Каким образом?

Д. Ф.: Это, конечно, нереалистичный срок, но, мне кажется, такие технологии вполне могут появиться в течение следующего десятилетия. Я думаю, что быстрый прогресс начнется в таких аспектах, как восприятие и контроль. Это сильно повлияет на общество, рынок труда, национальную безопасность и производительность. Третий аспект — понимание — при этом даже не будет затронут, но это позволит привлекать людей к решению задач, связанных с языком и мышлением.

Называя срок от трех до пяти лет, я имел в виду, что рабочие идеи у нас уже есть, просто нужны инвестиции в определенные разработки. Если бы я не знал способа достичь сильного ИИ, оценка срока была бы совсем другой. Пока деньги по большей части вкладываются в машинное обучение из-за быстрой окупаемости. Я же пытаюсь привлекать инвестиции в технологию, которая позволит развить понимание. В отличие от других, я не считаю, что для достижения сильного ИИ нужно ждать какого-то огромного прорыва. Думаю, мы уже знаем, как действовать, просто нужно доказать эффективность этого пути.

М. Ф.: То есть компания Elemental Cognition занимается вопросами сильного ИИ?

Д. Ф.: Мы работаем над созданием естественного интеллекта, способного автономно обучаться, читать и понимать, и таким способом мы хотим получить ИИ, умеющий свободно общаться с людьми.

М. Ф.: Мне известна всего одна компания, которая также занимается этой проблемой — DeepMind. Поразительно, насколько разные подходы вы применяете. В DeepMind ориентируются на

глубокое обучение с подкреплением в играх и симуляциях, а от вас я слышу, что путь к интеллекту лежит через язык.

Д. Ф.: Я бы немного переформулировал нашу цель. Мы хотим создать интеллект, основанный на логике, языке и разуме, совместимый с человеческим. Другими словами, мы работаем над системой, умеющей обрабатывать язык, как это делают люди. Помимо нейронных сетей мы применяем длинный диалог, формальные рассуждения и формальные логические представления. Для вещей, которые невозможно эффективно изучать с помощью нейронных сетей, мы находим другие способы работы.

М. Ф.: Если я правильно помню, еще вы работаете над обучением без учителя, которое, как мне кажется, повлияет на реальный прогресс в сфере ИИ.

Д. Ф.: Мы применяем оба варианта обучения. Например, используем обучение без учителя на больших текстах, а обучение с учителем — на аннотированных информационных материалах.

М. Ф.: Ждет ли нас в ближайшем будущем экономический спад с массовой потерей рабочих мест?

Д. Ф.: Не знаю, ждут ли нас такие же драматические события, как после промышленной революции, но мне кажется, что ИИ несет с собой не менее резкие перемены. Многие столкнутся с необходимостью переквалификации. Переход будет болезненным, но, в конце концов, он может привести к появлению еще большего числа рабочих мест. Потому что работы меньше не станет, изменится только ее характер.

М. Ф.: А как быть с проблемой несоответствия навыков?

Д. Ф.: Разумеется, новые рабочие места будут недоступны людям без специальной подготовки. Но я уверен, что появится и много других рабочих мест, например, в сфере здравоохранения, таких, в которых важен аспект общения с людьми.

Мы считаем, что в будущем нас ждет тесное и свободное партнерство людей и машинного интеллекта. Людям не нужно будет долго учиться, чтобы получить доступ к знаниям и эффективно их применять. Компьютеры же в рамках этого сотрудничества смогут лучше нас понять. Мы хотим, чтобы компьютеры стали совместимы с людьми, так почему бы не заплатить тем, кто поможет достичь этой цели? Экономика взаимодействия людей с машинами очень интересная тема, но нас ждет длинный переходный период.

М. Ф.: Как вы относитесь к опасностям, которые, если верить Илону Маску и Нику Бострому, несет в себе суперинтеллект?

Д. Ф.: Я думаю, что каждый раз, когда мы даем машинам какие-то рычаги влияния, появляется множество причин для беспокойства. Любая ошибка компьютера, контролирующего электрические сети, системы вооружения или беспилотные автомобили, может спровоцировать серьезную катастрофу. Ситуация становится еще более серьезной в случае проблем с компьютерной безопасностью. Об этом все время следует помнить. Тем более что с помощью компьютеров мы контролируем все больше и больше важных сфер, таких как транспорт, питание и национальная безопасность. Это пока еще не ИИ, но все подобные системы следует проектировать очень тщательно и продумывая возможные точки отказа.

Ник Бостром любит пугать публику, описывая, как у машины вдруг появляются какие-то собственные цели, для достижения которых она уничтожает человечество. Меня подобные прогнозы не беспокоят, потому что машину нужно сначала запрограммировать для такой реакции. Зачем ставить машину, предназначенную для производства скрепок, контролировать электрическую сеть? У нас слишком много более актуальных проблем, чтобы беспокоиться о гипотетической ситуации, когда у ИИ внезапно возникнут собственные желания и цели и он решит пожертвовать человеческой расой, чтобы сделать больше скрепок.

М. Ф.: Должно ли правительство регулировать сферу ИИ?

Д. Ф.: Раз уж машины могут принимать решения, влияющие на нашу жизнь, нужно четко определить, кто за что отвечает. Люди, на жизнь которых влияют принимаемые машинами решения, имеют право знать их мотивы. Правительство должно вмешиваться и прояснять, кто отвечает за происходящее и на что могут рассчитывать люди в случае правовых отношений с машиной.

Кроме того, нужно выработать критерии проектирования воздействующих на людей систем. Потому что в этом случае любые ошибки или взлом могут сильно повлиять на общество. Нужен вариант, не замедляющий развития технологий и одновременно контролирующий развертывание подобных систем. Требуется регулирование и на рынке труда.

М. Ф.: После вашего ухода компания IBM создала крупное бизнес-подразделение для работы с суперкомпьютером Watson, которое с переменным успехом пыталось извлекать прибыль из этого проекта. Что вы думаете об их опыте и проблемах, с которыми они столкнулись?

Д. Ф.: Я не знаю, что там происходит в настоящее время, но думаю, что если они хотели использовать суперкомпьютер Watson как бренд для входа в ИИ-индустрию, они получили такую возможность.

С точки зрения бизнеса положение IBM уникально. Они могут выйти на рынок, предлагая бизнес-аналитику, анализ данных и оптимизацию. И обеспечить целевое значение, например, в медицинских приложениях.

Трудно измерить, насколько успешными оказались их попытки получить прибыль как ИИ-компания. Оценка в данном случае зависит от бизнес-стратегии и от того, что вы считаете ИИ. Сейчас

в центре внимания потребителей виртуальные помощники Siri и Alexa. Насколько велика их ценность для бизнеса, я не знаю.

М. Ф.: Беспокоит ли вас преимущество в сфере ИИ, которое может получить Китай за счет количества открытых данных? Нужна ли промышленная политика для поддержания конкурентоспособности?

Д. Ф.: Я думаю, что гонка вооружений имеет место, и ее результаты повлияют на производительность, рынок труда, национальную безопасность и потребительские рынки. Чтобы оставаться конкурентоспособными, нужно инвестировать в ИИ, привлекать и поддерживать таланты.

И вот тут важно сохранить непростой баланс между мерами поддержания конкурентоспособности, мерами контроля и регулирования и вопросами конфиденциальности. Мне кажется, что для оптимального решения в такой ситуации нужны более вдумчивые и знающие лидеры. Чем выше будет их осведомленность в сфере ИИ, тем лучше. Может быть, для решения этих проблем нам потребуется ИИ!

М. Ф.: Ваши взгляды на будущее ИИ можно назвать оптимистичными?

Д. Ф.: По большому счету я оптимист. Думаю, что заниматься ИИ — наша судьба. Вспомните, что меня интересовало изначально. Компьютер дает возможность экспериментировать с самой природой интеллекта. Неразумно отказываться от этой возможности. Люди связывают самосознание с интеллектом. Так почему не сделать все от нас зависящее, чтобы лучше понять его сильные и слабые стороны и применить более эффективно?

Забавно, что человечество хочет исследовать космос, чтобы найти там другие виды интеллекта, хотя на самом деле такой интеллект растет рядом с нами. Я думаю, что в конечном итоге он поможет

нам повысить творческие способности и наш уровень жизни, применяя для этого способы, которых мы пока не можем себе представить.

Экзистенциальная угроза, конечно, существует, и я думаю, она изменит наше представление о себе и о том, что мы считаем уникальными качествами людей. И тут возникают очень интересные вопросы. Если любую задачу машины решат лучше, что произойдет с самооценкой людей? Куда переместится их самовыражение? В сторону эмпатии, эмоций, понимания и духовных вещей? Я не знаю, но искать ответ на эти вопросы рано или поздно придется. Мы просто не сможем этого избежать.

> *Пока что наши роботы не могут приблизиться по интеллекту даже к насекомым, поэтому появления суперинтеллекта в ближайшее время я не боюсь.*

Родни Брукс

ПРЕЗИДЕНТ КОМПАНИИ RETHINK ROBOTICS

Родни Брукс — всемирно известный специалист по робототехнике. Соучредитель компании iRobot — лидера в области производства потребительской робототехники и военных роботов.

Степень доктора computer science получил в Стэнфорде. В настоящее время — председатель и технический директор компании Rethink Robotics. Возглавлял Лабораторию ИИ, а затем — CSAIL.

Член нескольких организаций, включая AAAI, одним из основателей которой он является. Получил ряд наград за работу в области ИИ, в том числе награду Computers and Thought Award, техническую премию Инаба от IEEE за инновации, самую престижную в мире робототехники премию Эндельбергера и премию Robotics and Automation Award от IEEE.

Сыграл самого себя в фильме Эррола Мориса 1997 г. «Быстро, дешево и неуправляемо».

Мартин Форд: Еще работая в MIT, вы основали компанию iRobot. Сейчас это крупнейший дистрибьютор коммерческих роботов. Как это получилось?

Родни Брукс: iRobot я основал в 1990 г. совместно с Колином Энглом и Хелен Грейнер. Первые 14 бизнес-моделей оказались неудачными, но в 2002 г. мы нашли пару работающих вариантов. Во-первых, мы решили производить роботов для военных. Эти роботы применялись в Афганистане для осмотра пещер. Позднее во время афганских и иракских конфликтов примерно 6500 наших роботов использовались в качестве саперов.

Тогда же, в 2002 г., мы начали выпуск робота-пылесоса Roomba. К настоящему моменту продано уже более 20 млн единиц. А в 2017 г. годовой доход компании составил 884 млн долларов. Можно сказать, что это самый успешный робот-пылесос из когда-либо поставляемых на рынок. В его основе лежит интеллект насекомого, разработкой которого я начал заниматься еще в 1984 г. в MIT.

В 2010 г. я покинул MIT и основал компанию Rethink Robotics, создающую промышленных роботов. На сегодняшний день они используются на тысячах предприятий. В отличие от множества других роботов, они совершенно безопасны и их не нужно изолировать. Более того, программное обеспечение Intera 5 позволяет роботу написать программу, реализующую то, что ему показали. Это графическая программа, которая выводит деревья поведения. При желании пользователь может ими манипулировать, внося в предложенный роботом план свои коррективы. Наши роботы используют тактильную обратную связь и зрение и работают по всему миру круглые сутки.

М. Ф.: Как возник ваш интерес к робототехнике и ИИ?

Р. Б.: Мое детство прошло в Аделаиде, в Южной Австралии. В 1962 г. мама нашла две американские детские энциклопедии из серии «Что есть что». Одна называлась «Электричество», дру-

гая — «Роботы и электронный мозг». Остаток детства я провел, используя сведения из этих энциклопедий для исследований и попыток создать интеллектуальный компьютер и, в конечном итоге, робота.

В Австралии я получил степень бакалавра математики и начал работать над кандидатской диссертацией по ИИ, но понял, что в стране нет отделов computer science. Я подал документы в три вуза, в которых, как я слышал, проводились исследования в области ИИ. Из MIT пришел отказ, а вот Университет Карнеги — Меллона и Стэнфорд были готовы меня принять. Я выбрал Стэнфорд, потому что он ближе к Австралии.

В 1977 г. началась моя работа над диссертацией по компьютерному зрению под руководством Тома Бинфорда. Затем была научно-исследовательская деятельность в Университете Карнеги — Меллона и работа в MIT, пока в 1983 г. я не вернулся в Стэнфорд в качестве кандидата на штатную должность в профессорско-преподавательском составе. Через год я стал преподавателем в MIT, где проработал 26 лет.

Именно во время первой стажировки в MIT я начал работать над интеллектуальными роботами. К моменту моего возвращения в 1984 г. я понял, как мало достигнуто в моделировании восприятия роботов. Меня вдохновили насекомые, которые со своими сотнями тысяч нейронов превосходили любого из наших роботов. Я взял их интеллект за основу для своей модели, которой занимался в течение нескольких лет.

Затем я руководил основанной Марвином Минским Лабораторией ИИ. Со временем ее объединили с Лабораторией computer science, сформировав CSAIL.

М. Ф.: Каким из своих достижений вы гордитесь больше всего?

Р. Б.: В марте 2011 г. в Японии случилось землетрясение, и приливная волна вывела из строя АЭС «Фукусима». Примерно через

неделю мы узнали, что на территорию станции не могут доставить роботов, которые должны произвести осмотр и выяснить, что произошло. За 48 часов из iRobot доставили шесть роботов и обучили местную техническую команду ими управлять. Позднее мы узнали, что именно благодаря нашим роботам удалось остановить реакторы.

М. Ф.: Я помню эту историю. Я тогда сильно удивился, ведь считалось, что в Японии самые передовые разработки в области робототехники.

Р. Б.: Тогда мы все получили наглядный урок. Дело в том, что пресса представляла японские разработки более совершенными, чем они есть на самом деле.

Тысячи наших роботов в течение девяти лет применялись в зонах боевых действий. Красивым внешним видом они похвастаться не могли, функциональность на базе ИИ практически отсутствовала, но они реально работали. Большую часть своей жизни я объяснял людям, как они заблуждаются, когда после просмотра видеороликов думают, что вот-вот начнутся крупные перемены, потому что роботы займут все рабочие места и наступит массовая безработица.

Если 30 лет назад в компании Rethink Robotics не было демонстрационных образцов; наивно думать, что уже завтра мы сможем показать готовый продукт. Переход от моделей, работающих в лабораторных условиях, к реальному продукту занимает много времени. Сейчас публика восторгается беспилотными автомобилями. Но мало кто помнит, что первый автомобиль, который автономно проехал по автостраде 10 миль со скоростью более 55 миль в час, появился в 1987 г. недалеко от Мюнхена. Автомобиль, в котором водитель пользовался только педалями и совсем не трогал руль, в первый раз проехал от одного побережья США до другого в 1995 г. в рамках проекта No Hands Across America. Означает ли это, что завтра мы увидим серийные беспилотные автомобили?

Нет, для их разработки требуется много времени, а люди переоценивают скорость внедрения подобных технологий.

М. Ф.: То есть вы не согласны с законом ускоряющейся отдачи Курцвейла? Вы не верите в постоянное ускорение прогресса?

Р. Б.: Глубокое обучение дает фантастические результаты, и люди со стороны искренне воспринимают это как чудо. Разговоры об экспоненциальном росте начались с закона Мура, но постепенно этот закон перестает работать, потому что нельзя до бесконечности делить пополам место на кристаллах интегральных схем. Зато постепенно возобновляются работы над компьютерной архитектурой. В течение 50 лет никто не мог позволить себе делать что-то необычное, потому что его сразу же обогнали бы конкуренты. А сейчас мы вступаем в эру компьютерной архитектуры.

Заключительная часть вышедшей в 1965 г. статьи Мура *The Future of Integrated Electronics* («Будущее интегрированной электроники») посвящена случаям, в которых его закон неприменим. Например, он не работает при аккумулировании энергии.

Десять лет назад на этом обожглись венчурные инвесторы из Кремниевой долины, которые были уверены, что закон Мура работает везде, в том числе и в зеленых технологиях. Но зеленые технологии опираются на объем и энергию. Это не та ситуация, когда на физически вдвое меньшем пространстве все равно можно хранить все тот же объем информации.

Фундаментальный для глубокого обучения метод обратного распространения разработан еще в 1980 гг. А фантастические результаты с его помощью были получены через 30 лет. А тогда, в 1980-х гг., от этого метода хотели отказаться, потому что он ничего не давал. По этой причине отказывались от сотен вещей. Но случилось так, что в комбинации с ограничениями на значения весов, увеличением числа слоев и ростом вычислительных ресурсов метод обратного распространения дал интересные результаты.

Тем не менее успехи глубокого обучения не могут продолжаться вечно. Все имеет предел. И Рэю Курцвейлу не придется в ближайшее время загружать свое сознание в компьютер. Биологические системы работают по другим принципам. Они опираются на сотни алгоритмов. И чтобы добиться прогресса в этой области, нужно создать эти сотни алгоритмов. Поэтому при каждой встрече с Курцвейлом я напоминаю ему, что он рано или поздно умрет.

М. Ф.: Это очень грубо.

Р. Б.: Я тоже умру. Только я в этом не сомневаюсь, а он принадлежит к людям, которые превращают технологию в религию. Миллиардеры из Кремниевой долины финансируют компании, дающие возможность загрузить свое сознание в машину, как обещает Рэй Курцвейл. Но я уверен, что еще много веков мы будем оставаться обычными смертными.

М. Ф.: С этим не поспоришь. А как быстро на рынок могут выйти беспилотные автомобили? Говорят, что компания Google уже запустила первые образцы в Аризоне.

Р. Б.: Подробностей я пока не знаю, но этот процесс тоже займет намного больше времени, чем кажется на первый взгляд. Города Маунтин-Вью (Калифорния) и Феникс (Аризона) сильно отличаются от других городов США. Там могут испытываться демонстрационные версии, но до появления реальной услуги, приносящей прибыль, пройдет минимум несколько лет. Под прибылью я имею в виду заработок почти с той же скоростью, с которой Uber теряет деньги. В прошлом году эта сумма составила 4,5 млрд долларов.

М. Ф.: То есть вы хотите сказать, что поскольку Uber теряет деньги на каждой поездке, переход к беспилотным автомобилям — это нерациональная бизнес-модель?

Р. Б.: Сегодня утром я читал, что средняя почасовая заработная плата водителя Uber составляет 3,37 доллара. Это не такая большая

сумма, чтобы ради ее экономии заменять водителей дорогими датчиками. При этом у нас пока нет реальной модели беспилотного автомобиля. Компания Google ставит на крышу своим машинам кучу дорогих датчиков, Tesla попробовала использовать встроенные камеры и тоже потерпела неудачу. Нет, мы можем посмотреть впечатляющие демонстрации, но все это заранее подготовлено. Как и видеоролики с японскими роботами.

М. Ф.: Вы сейчас говорите о фальсификации?

Р. Б.: Нет. Просто публике показывают красивую картинку, оставляя за кадром множество деталей. Люди самостоятельно дорисовывают скрытое, делают обобщения и приходят к выводам, которые не имеют отношения к реальности.

Например, в отличие от Феникса, в котором проводилась демонстрация, улицы Кембриджа, штат Массачусетс, в котором я живу, забиты машинами. И, как правило, движение на них одностороннее. Я не могу даже представить, как осуществить посадку в беспилотный автомобиль у меня в районе. Он подберет пассажира посреди дороги? Или займет автобусную полосу? Обычно ожидающее пассажира такси блокирует дорогу, значит, беспилотный автомобиль должен быть быстрым, чтобы не вызывать возмущения остальных участников дорожного движения. Мне кажется, что даже в Фениксе долгое время посадка и высадка будет осуществляться только в специально выделенных местах. Беспилотные автомобили не могут просто встроиться в существующую дорожную сеть.

Компания Uber уже начала создавать специальные места для своих такси. Новую систему пробовали в Сан-Франциско и Бостоне, и теперь она применяется уже в шести городах, где в любую погоду приходится стоять в очереди в ожидании своего такси.

Люди думают, что беспилотные автомобили отличаются от обычных только отсутствием водителя. На самом деле для перехода

к ним нужны серьезные преобразования. Вспомните, как изменились наши города после появления автомобилей.

М. Ф.: Сколько времени может потребоваться, чтобы получить беспилотный эквивалент Uber?

Р. Б.: Это будет постепенный процесс. Сначала, скорее всего, пассажирам придется идти в указанную точку. Как при аренде автомобилей от Zipcar, которые стоят на специально выделенных для них парковочных местах. Я не знаю, увижу ли я при жизни множество беспилотных автомобилей, курсирующих по городам. Для зарядки беспилотных автомобилей нужно создать соответствующую инфраструктуру.

М. Ф.: Мне приходилось слышать, что беспилотное такси появится лет через пять. Вы не согласны?

Р. Б.: Да. Возможно, будут реализованы некоторые аспекты этого сервиса, но до эквивалента обычных такси еще далеко. Для начала потребуется множество новых компаний, которые будут заниматься поддержкой. Скорее всего, пассажир будет общаться с машиной голосом. Кроме того, требуется система регулирования. На какие команды пассажира будет реагировать машина? Будет ли она принимать сообщения от ребенка? Сейчас вы можете посадить в такси 12-летнего мальчика и попросить доставить его по указанному адресу. С беспилотными автомобилями это, скорее всего, станет невозможным.

М. Ф.: Давайте вернемся к интеллекту насекомых. Я всегда считал, что они — хорошо запрограммированные биологические роботы. Приближаются ли роботы к тому, на что способно насекомое? И приближает ли это нас к суперинтеллекту?

Р. Б.: Пока что наши роботы не могут приблизиться по интеллекту даже к насекомым, поэтому появления суперинтеллекта в ближайшее время я не боюсь. Мы не можем воспроизвести способности насекомых к обучению и адаптации и повторить их навыки.

М. Ф.: Вспомните 1990-е гг., когда создавалась компания iRobot. Оправдал ли прогресс в области робототехники ваши ожидания или вы разочарованы?

Р. Б.: Когда в 1977 г. я приехал в США, меня действительно интересовали роботы, но в итоге я занялся компьютерным зрением. На тот момент в мире существовало три мобильных робота. Один в Стэнфорде, где Ханс Моравек заставил его пройти 18 метров за 6 часов, второй — в Лаборатории реактивного движения НАСА (JPL), а третий — в Лаборатории анализа и архитектуры систем (LAAS) в Тулузе, Франция.

Сейчас компания iRobot выпускает миллионы мобильных роботов в год, и я счастлив видеть, как далеко все зашло. Наши достижения не выглядят как что-то особенное только потому, что за тот же период времени произошел переход от мейнфреймов размером с комнату к миллиардам смартфонов.

М. Ф.: Я знаю, что вы работали не только над интеллектом насекомых, но и над созданием роботизированных рук. Какой прогресс достигнут в этой области?

Р. Б.: Да, я хотел разграничить работу над мобильными коммерческими роботами в iRobot и занятия с моими студентами, поэтому в MIT я переключился на человекоподобных существ и руки. Работа продвигается медленно. Есть достижения, но это всегда решение каких-то узких задач.

М. Ф.: Проблемы в аппаратном или программном обеспечении? В конструкции или управлении?

Р. Б.: Это целый ряд параллельно идущих проблем. Нужен одновременный прогресс в конструкции, материалах поверхности, встроенных датчиках и алгоритмах управления ими. В видеороликах, демонстрирующих новые роботизированные руки, зачастую их держит человек. И по факту все, что там показывается, можно было бы с таким же успехом сделать с помощью маленькой пла-

стиковой игрушки с манипулятором. То есть у нас отсутствует ключевой компонент.

М. Ф.: Я видел отчеты, согласно которым с помощью глубокого обучения и обучения с подкреплением роботы осваивают новые навыки, практикуясь или даже просто просматривая обучающие видео на YouTube.

Р. Б.: Нам показывают лабораторные демонстрации. В компании DeepMind есть группа, работающая с нашими роботами. Недавно они опубликовали несколько интересных работ по тактильной обратной связи, где роботы прикрепляют зажимы к различным вещам. Это результат многомесячной работы команды исследователей. При этом любой здоровый человек, которому один раз показали, как это делается, сразу же справится с задачей. Роботы и близко не показывают подобных результатов.

Люди говорят, что сборка мебели из IKEA — отличный тест для роботов. Когда я делал это у себя дома, мне пришлось выполнить примерно 200 различных операций. Если взять роботов нашего производства, придется плотно работать несколько месяцев, чтобы получить грубую демонстрацию одной из этих 200 задач. Вот так обстоят дела в реальности.

М. Ф.: А можем ли мы в ближайшее десятилетие ожидать прорыва в области робототехники и ИИ?

Р. Б.: Прорывы совершаются неожиданно. Глубокое обучение — замечательная технология. Она позволила нам сделать фантастический шаг вперед и начать использовать речевые системы для Amazon Echo и Google Home. Его потенциал этим не исчерпывается, но все равно что-то придет ему на смену лет через десять.

М. Ф.: Под глубоким обучением вы подразумеваете нейронные сети, использующие метод обратного распространения ошибки?

Р. Б.: Да, нейронные сети со множеством слоев.

М. Ф.: Может, на смену придут нейронные сети, но с другим алгоритмом или байесовские сети?

Р. Б.: Возможно. Но я гарантирую, что в течение 10 лет появится новое направление, которое приведет к появлению других технологий. Я пока не знаю, какими они будут, но могу сказать о влиянии рынка, основой для которого послужат появляющиеся уже сейчас глобальные тенденции.

Например, резко меняется соотношение пожилых людей и людей трудоспособного возраста. В разных источниках встречаются цифры от 9 : 1 до 2 : 1. Количество пожилых людей в мире растет. И рынок будет стремиться помогать им. В Японии на выставках робототехники мы уже видели множество лабораторных демонстраций роботов, помогающих выполнять простые задачи, например, вставать с кровати, заходить в ванную и другие повседневные вещи. Сейчас этим занимаются сиделки, но так как доля трудоспособного населения сокращается, в какой-то момент для удовлетворения этой потребности просто не окажется рабочей силы. А значит, начнется производство роботов для помощи пожилым людям.

М. Ф.: Сегмент ухода за пожилыми людьми — действительно огромная возможность для индустрии робототехники и ИИ, но для такой деятельности необходима большая ловкость.

Р. Б.: Разумеется, сиделку нельзя просто взять и заменить роботизированной системой. Но спрос на такие вещи будет огромным, а значит, мотивированные люди начнут искать решение.

Кроме того, мне кажется, что увеличится спрос на строительные работы, потому что урбанизация идет с невероятной скоростью. Мы же до сих пор используем в строительстве множество методик, изобретенных еще римлянами. Тут явно есть место для технологических новшеств.

М. Ф.: Нас ждут роботы-строители или масштабная 3D-печать?

Р. Б.: 3D-печать может пригодиться. Никто не будет печатать целые здания, но вполне возможно печатать предварительно сформированные компоненты. Гораздо больше деталей будет изготавливаться вне стройплощадок, что, в свою очередь, приведет к инновациям в доставке, подъеме и перемещении.

Уже пошли разговоры о городском фермерстве и о переводе сельского хозяйства с полей на фабрики. Здесь может пригодиться машинное обучение. Современные вычислительные мощности позволяют создать замкнутый цикл, обеспечивающий семена питательными веществами и условиями, в которых они нуждаются, и выращивать еду, не беспокоясь о погоде на улице. Я думаю, что изменение климата приведет к новым вариантам автоматизации сельского хозяйства.

М. Ф.: А как насчет бытовых роботов? Тех самых, которые приносят вам пиво.

Р. Б.: Генеральный директор компании iRobot Колин Энгл говорит об этом уже 28 лет. Но нам еще долго придется ходить за пивом самостоятельно.

М. Ф.: Может ли когда-либо появиться действительно универсальный потребительский робот, который будет делать жизненно необходимые вещи?

Р. Б.: Робот-пылесос Roomba можно считать незаменимым? Нет. Но он приносит пользу при достаточно низких затратах. Хотя речь идет не о необходимости, а об уровне удобства.

М. Ф.: Когда появится робот, который может не только перемещаться и пылесосить полы?

Р. Б.: Если бы я это знал! И никто не знает. Публику пугают роботами, захватывающими мир, а мы даже не можем ответить на вопрос, когда они научатся приносить пиво.

М. Ф.: Я недавно читал статью, в которой исполнительный директор компании Boeing Деннис Мюленбург говорит, что в следующем десятилетии людей будут перевозить автономные такси-дроны.

Р. Б.: С таким же успехом можно сказать, что у нас будут летающие машины. Бывший генеральный директор компании Uber Трэвис Каланик объявлял, что к 2020 г. у них появятся летающие такси. Но ничего подобного явно не произойдет. Это не значит, что автономный личный транспорт невозможен в принципе. У нас уже есть вертолеты и другие машины, которые могут надежно перемещаться с места на место без пилота. Их выход на массовый рынок зависит от экономики.

М. Ф.: Достижим ли сильный ИИ? И когда, по вашим оценкам, мы приблизимся к его созданию?

Р. Б.: Я считаю сильный ИИ возможным. И по моим оценкам, это будет 2200 г., но я могу ошибаться.

М. Ф.: Какие препятствия нужно для этого преодолеть?

Р. Б.: Я уже упоминал о недостатке ловкости. Для понимания мира важна способность ориентироваться в нем и манипулировать объектами, при этом физическим контекстом дело не ограничивается. Например, ни один робот или ИИ-система не знает, чем, кроме цифры в календаре, сегодняшний день отличается от вчерашнего. У них нет воспоминаний о личном опыте, нет понимания долгосрочных целей и постепенного продвижения к ним. Любая современная программа ИИ напоминает слабоумного гения, живущего исключительно настоящим. А раз сильный ИИ эквивалентен человеку, он должен полностью осознавать эти вещи.

В 1980-х гг. целое сообщество занималось моделированием адаптивного поведения. Я тоже принимал в этом участие. Особого прогресса с тех пор нет, мы даже не можем указать, в какую сто-

рону следует двигаться. Сейчас над этим никто не работает, хотя есть люди, утверждающие, что они продвигают сильный ИИ. Мы думаем, что мы особенные и живем в золотые времена. Но так думали и многие до нас.

М. Ф.: Существуют опасения, что Китай обгонит нас в сфере ИИ благодаря более многочисленному населению и, как следствие, большему количеству данных, а также отсутствию строгих правил, касающихся конфиденциальности. Как вы думаете, вступаем ли мы в новую гонку вооружений?

Р. Б.: Соревнование в этой сфере неизбежно. Оно было между компаниями и будет между странами.

М. Ф.: Опасно ли для стран Запада, если лидером в этой сфере окажется такая страна, как Китай?

Р. Б.: Я не думаю, что все так просто. Мы увидим неравномерное использование ИИ-технологий. Например, в Китае уже сейчас лица распознаются способом, неприемлемым в США. Что касается процессоров AI chip, здесь отставание непозволительно. Но для поддержания темпа нужны ресурсы.

Между тем политики говорят, что нам нужно больше работников угольной промышленности. Научные же бюджеты сокращаются, меньше финансируются даже такие места, как Национальный институт стандартов и технологий. Эти отсталые взгляды приносят вред.

М. Ф.: Но ведь ИИ и робототехника несут в себе не только благо. Многие считают, что мы на пороге новой промышленной революции. Как вы оцениваете влияние новых технологий на рынок труда и экономику?

Р. Б.: Влияет не сам ИИ, а то, как его используют. Влияет оцифровка мира и создание новых решений. Например, в США на большей части платных дорог и мостов оплата принимается авто-

матически, а раньше это делали люди. На лобовое стекло теперь ставится метка с цифровой подписью автомобиля, а ИИ-система с глубоким обучением позволяет надежно ее прочитать. Это достижение в сфере компьютерного зрения позволило избавиться от сборщиков платы за проезд. Причем при оплате можно обойтись даже без физического предоставления кредитной карты, и не требуются автомобили, чтобы доставлять деньги в банк, когда есть цифровая логистическая цепь.

Как видите, для автоматизации работы сборщика оплаты за проезд потребовался целый набор цифровых компонентов. ИИ всего лишь один из них. И замена человека на ИИ-систему произошла не за одну ночь. Это постепенный процесс, который медленно модифицирует рынки труда.

М. Ф.: Цифровые сети закроют много вакансий, не требующих высокой квалификации?

Р. Б.: Есть ряд недооцененных вещей, необходимых для поддержания нашего общества, таких как помощь пожилым людям и инвалидам. Школьные учителя лишены признания и достойной заработной платы. И я не знаю, как сделать эту работу экономически выгодной. Раз часть рабочих мест исчезнет из-за автоматизации, как мы можем распознать и оценить по достоинству рабочие места, которым такая опасность не угрожает?

М. Ф.: То есть вы считаете, что массовая безработица нам не опасна. Но мне кажется, что исчезнет много престижных рабочих мест. Офисный сотрудник с высшим образованием окажется не у дел, в то время как горничная в отеле продолжит спокойно работать.

Р. Б.: Может быть и такое. Но причина будет не в ИИ, а в массовом переходе к цифровым решениям.

М. Ф.: Но ведь именно на этой платформе будет развернут ИИ, что приведет к ускорению процесса.

Р. Б.: Да, при наличии подходящей платформы развернуть ИИ проще. Но эта платформа базируется на компонентах, которые могут быть взломаны кем угодно.

М. Ф.: Раз речь зашла о безопасности, о чем еще имеет смысл беспокоиться?

Р. Б.: Основной вопрос связан именно с безопасностью. Меня волнует отказ от безопасности и конфиденциальности цифровых цепочек, который мы добровольно совершили в обмен на простоту использования технологий. Мы уже видели применение социальных сетей в качестве оружия. Вероятнее всего, мы увидим агрессию не от ИИ, а от людей, использующих его.

М. Ф.: А что вы думаете по поводу вооружения роботов и дронов? Стюарт Рассел снял об этом довольно страшный фильм «Роботы-убийцы».

Р. Б.: Сегодня такие вещи вполне возможны, ИИ для них не нужен. Фильм был импульсивной попыткой сказать, что роботы и война — плохая комбинация. Я считаю немного по-другому. Робот может дожидаться, пока в него выстрелят, и стрелять только в ответ. Парень 19 лет в чужой стране в темное время суток, когда вокруг полно вооруженных людей, такой роскоши себе позволить не может.

Говорят, что проблему можно решить, запретив применение ИИ в армии. На самом деле нужно понять, чего мы не хотим, определиться с проблемой, которая не связана с ИИ. Скажем, наш следующий полет на Луну будет зависеть от ИИ и машинного обучения, но в 1960-х гг. мы справились и без этого.

М. Ф.: Что вы думаете о проблеме контроля сильного ИИ и о комментариях Илона Маска?

Р. Б.: Когда в 1789 г. жители Парижа впервые увидели воздушный шар, они были обеспокоены тем, что там, наверху, из людей высасывают души.

Я написал эссе *The Seven Deadly Sins of Predicting the Future of AI*[1] («Семь смертных грехов в предсказании будущего ИИ»), и все эти грехи крутятся вокруг вышесказанного. Люди представляют точно такой же мир, как сейчас, но с суперразумом в центре. На самом же деле нас ждет видоизмененный мир.

М. Ф.: Нужно ли будет регулировать технологические прорывы, когда они произойдут?

Р. Б.: Как я уже говорил, регулировать нужно действия, а не инструменты. Должны ли мы прекратить исследования на оптических компьютерах, позволяющих намного быстрее выполнять матричное умножение и быстрее применять глубокое обучение? Нет, это безумие. Будет ли разрешено беспилотным грузовым автомобилям становиться на длительную стоянку в переполненных районах Сан-Франциско? Это уже нужно регулировать. А технологии лучше оставить в покое.

М. Ф.: Учитывая все вышесказанное, полагаю, вас можно назвать оптимистом? Вы верите, что преимущества всего этого перевесят любые риски?

Р. Б.: Конечно. Мир уже перенаселен, и чтобы выжить, мы должны идти этим путем. Меня сильно беспокоит снижение уровня жизни, потому что по мере старения населения перестанет хватать рабочей силы. Меня волнуют вопросы безопасности и конфиденциальности.

Голливудская идея владычества сильного ИИ уже давно захватила умы, и мы тратим ресурсы на решение неактуальной проблемы, в то время как нужно беспокоиться о сегодняшних опасностях и рисках.

[1] https://rodneybrooks.com/the-seven-deadly-sins-of-predicting-the-future-of-ai/

> *Меня не так пугает гипотетическое порабощение человечества сверхинтеллектом, как люди, которые могут злоупотребить новыми технологиями.*

Синтия Бризил

ДОЦЕНТ КАФЕДРЫ МЕДИАИСКУССТВА И НАУК, ДИРЕКТОР ГРУППЫ ПЕРСОНАЛЬНЫХ РОБОТОВ В МЕДИАЛАБОРАТОРИИ MIT, СМОТРИТЕЛЬ МУЗЕЯ НАУКИ В БОСТОНЕ, ОСНОВАТЕЛЬ КОМПАНИИ JIBO, INC.

Синтия Бризил — создатель первого в мире социального робота Kismet. Многие из известных разработок — от миниатюрных роботов-гексаподов и до роботизированных технологий, встроенных в повседневные вещи, — принадлежат Синтии. Журнал Time включил ее в список лучших изобретателей 2008 и 2017 гг. и признал ее робота Jibo одним из лучших изобретений 2017 г. В 2014 г. признана самой перспективной женщиной-предпринимателем по версии журнала Fortune.

Автор книги Designing Sociable Robots («Разработка социальных роботов») и более 200 статей на темы социальной робототехники, взаимодействия людей с роботами, автономной робототехники, ИИ и обучения роботов. Лауреат премии Gilbreth Lecture Национальной инженерной академии, премии Джорджа Р. Стибитца в области компьютерных и коммуникационных технологий за «выдающийся вклад в развитие социальной робототехники и взаимодействия человека с роботом» и премии для молодых исследователей от Управления военно-морских исследований. Финалистка Национальной премии от Смитсоновского музея дизайна в области коммуникаций.

Мартин Форд: Когда, по вашим ощущениям, персональные роботы станут настоящим продуктом массового потребления, таким же, как телевизоры и смартфоны?

Синтия Бризил: Мне кажется, это время уже пришло. В 2014 г., когда я собирала средства для стартапа Jibo, все думали, что нам предстоит конкурировать со смартфонами. Ведь люди активно пользовались именно этой технологией, чтобы взаимодействовать с окружающим миром и контролировать его через сенсорный экран. На Рождество компания Amazon анонсировала виртуального помощника Alexa. Это открыло множество новых возможностей, потому что мы увидели, с какой охотой люди используют голосовые устройства.

В 2014 г. большинство тех, кто взаимодействовал с ИИ на потребительском уровне, составляли пользователи Siri или Google Assistant. Прошло всего четыре года, и многие перешли на устройства, оснащенные ИИ. Но мы находимся в самом начале эпохи взаимодействия компьютеров и людей. Данные, которые мы собрали с помощью робота Jibo, ясно показывают, что глубокое социально-эмоциональное, персонализированное, активное участие ИИ-систем позволяет поддержать людей в их переживаниях.

Сначала появились системы с голосовым интерфейсом, которые получали информацию о погоде или новости, но постепенно эти устройства осуществили возможность учиться, не выходя из дома, получить медицинскую консультацию без поездки к врачу и т. д.

М. Ф.: Но люди беспокоятся, например, о том, что активное взаимодействие с виртуальным помощником может повлиять на детское развитие. А роботов-компаньонов для пожилых людей рассматривают как антиутопию. Как вы к этому относитесь?

С. Б.: Скажу только, что есть наука, которую мы двигаем вперед, и машины, поддерживающие человеческие ценности, которые

еще не созданы. Критика есть всегда, и это простой путь. Но сейчас мы живем в обществе, где многие не могут получить помощь. А наши технологии позволяют обеспечить масштабируемую, доступную, эффективную и персонализированную поддержку. Мне кажется неправильным упускать эту возможность.

М. Ф.: Каким вы видите будущее Jibo? Превратится ли он в робота, который бегает по дому, выполняя полезную работу, или же он будет больше фокусироваться на социальных вещах?

С. Б.: Я думаю, что со временем появится множество разных видов роботов. Jibo — первый в своем роде. Его создавали как платформу с расширяемыми навыками без очерченной специализации. В компании Toyota Research Institute разрабатывают мобильных роботов для поддержки пожилых людей. При этом исследователи признают, что такие роботы должны обладать социальными и эмоциональными навыками.

А дальше все будет зависеть от ценностного предложения. Понятно, что одинокому старику и родителям, которые хотят, чтобы их ребенок выучил второй язык, потребуются разные роботы. Будут учитываться роль робота в доме и другие факторы, такие как цена. Этот сегмент постепенно будет расти и расширяться, и эти системы появятся в частных домах, школах, больницах и различных учреждениях.

М. Ф.: Как возник ваш интерес к робототехнике?

С. Б.: Я выросла в Ливерморе, штат Калифорния, где есть две национальные лаборатории. Мои родители работали там в качестве специалистов по computer science, так что техническое проектирование и computer science считались в нашем доме областями, дающими отличные карьерные возможности. Мне рано подарили Lego, потому что родители ценили такие игрушки.

В те годы у детей было не так много возможностей для работы с компьютерами, как сейчас. Но я посещала Национальную ла-

бораторию, где проводились занятия для детей. Я помню перфокарты! Благодаря родителям я получила доступ к компьютерам намного раньше большинства моих сверстников. Неудивительно, что наш дом был одним из первых, в котором появились персональные компьютеры.

Когда мне было 10 лет, на экраны вышел первый фильм «Звездные войны», и меня очаровали роботы, обладающие эмоциями и вступающие в отношения друг с другом и с людьми.

М. Ф.: Вашим научным руководителем был Родни Брукс. Как он повлиял на вашу карьеру?

С. Б.: Я хотела стать космонавтом-исследователем. Для этого была нужна докторская степень в соответствующей области. Я выбрала космическую робототехнику и подала заявления в аспирантуры нескольких вузов. На дне открытых дверей в MIT меня впечатлила лаборатория мобильных роботов Родни Брукса, которые имитировали насекомых. Подход к интеллекту с точки зрения биологии был нетипичным для этой области. Темой моей работы стало пересечение вещей, которые можно извлечь из этологии и других форм интеллектуальной деятельности, и машинного интеллекта.

Родни Брукс в то время работал над маленькими движущимися роботами и написал статью *Fast, cheap and out of control: a robot invasion of the Solar cystem* («Быстро, дешево и неуправляемо: вторжение роботов в Солнечную систему»). Там он выступал за то, чтобы вместо одного или двух очень больших и очень дорогих самоходных аппаратов посылать на Марс много-много маленьких. Это была очень важная статья, и моя магистерская диссертация заключалась в разработке первых планетарных роботов наподобие Micro Rover. Я получила возможность работать в Лаборатории реактивного движения (JPL). Мне нравится думать, что некоторые из моих исследований внесли вклад в создание марсохода Sojourner и космическую программу Pathfinder.

Я уже заканчивала работу над магистерской диссертацией, когда Родни взял творческий отпуск. После возвращения он объявил, что теперь мы будем делать антропоморфных роботов. Это стало для всех полной неожиданностью, потому что считалось, что от насекомых мы перейдем к рептилиям и, возможно, к млекопитающим, то есть будем развивать эволюционную цепочку интеллекта. Но Родни настаивал на своем, так как он увидел, что в Азии, особенно в Японии, уже начали разрабатывать антропоморфных роботов. Будучи старшим аспирантом, я руководила разработками, направленными на изучение теории воплощенного познания. Согласно этой теории, физическое тело влияет на природу интеллекта, который может иметь или научиться развивать машина.

Когда 5 июля 1997 г. NASA в рамках проекта Mars Pathfinder опустило марсоход Sojourner на поверхность Марса, я задумалась, что мы разрабатываем удивительных автономных роботов только для экспертов, а не для всех людей.

Мы уже знали, что люди наделяют автономных роботов человеческими качествами, пытаясь понять их, поэтому мы предположили, что универсальность нам обеспечит социальный межличностный интерфейс. До этого исследователей интересовало, как интеллектуальные машины манипулируют неодушевленными предметами. Мы же начали думать об их сотрудничестве и общении с людьми естественным для нас образом.

Большая часть моей докторской диссертации к тому моменту уже была готова, но я вошла в кабинет Рода и сказала: «Нужно все переписать. Диссертацию следует посвятить роботам в жизни обычных людей, социально и эмоционально разумным роботам». Надо отдать Родни должное: он понял, что это станет ключом к интеграции роботов в повседневную жизнь. И он позволил мне пойти на это. В результате появился первый в мире социальный робот Kismet.

М. Ф.: Я знаю, что сейчас он находится в музее MIT.

С. Б.: Робот Kismet действительно положил начало всему. Он напоминал дроидов из «Звездных войн». Я понимала, что не смогу создать автономного робота, который мог бы конкурировать с социальным и эмоциональным интеллектом взрослых людей. Я решила рассмотреть отношения между взрослым и младенцем, чтобы понять, как зарождается коммуникабельность и как она развивается со временем. Робот Kismet моделировал невербальное, эмоциональное общение младенца.

Ведь способность формировать эмоциональную связь — часть нашего механизма выживания. Именно она заставляет родителей или других взрослых людей относиться к новорожденному или маленькому ребенку как к полноценной социальной единице. На этой стадии у ребенка начинает развиваться социальный и эмоциональный интеллект. Фактически идет процесс начальной загрузки. Но эти способности не развиваются, если ребенок растет в неправильной социальной среде.

В рамках этой философии Kismet был смоделирован с внешностью птенца. Я тогда читала много литературы, посвященной анимации. Там рассматривались способы создания вещей, вызывающих у людей эмоциональный отклик и желание заботиться. Мне хотелось, чтобы люди могли взаимодействовать с моим роботом на уровне подсознания и развивать его естественным образом. Робот был спроектирован так, чтобы все его характеристики (движения, издаваемые им звуки и его внешний вид) помогали формировать вокруг него правильную социальную среду, дающую ему возможность учиться и развиваться. Невербальная коммуникация составляет огромную часть нашего общения и служит основой для множества социальных суждений.

Взаимодействие с современными голосовыми помощниками очень транзакционно. Оно напоминает игру в шахматы. В литературе по психологии часто встречается термин «социальные

танцы», который представляет процесс общения как взаимную адаптацию в танце. Невербальные сигналы влияют на выводы, сделанные собеседниками в динамично связанном дуэте. Встроить в машины социальный и эмоциональный интеллект крайне трудно.

М. Ф.: А также требует огромных вычислительных ресурсов?

С. Б.: Правильно. Причем даже бóльших ресурсов, чем зрение или манипулирование. Интеллект и поведение машины должны совпадать с человеческими в умении делать выводы и прогнозировании наших мыслей, намерений, убеждений, желаний. Мне хотелось создать машину, которая сможет оказывать помощь и поддержку в социальной и эмоциональной сфере. Мы начали искать новые приложения для роботов в областях, связанных с трансформацией человека, таких как образование, изменение поведения, улучшение самочувствия, тренировка… Об этих вещах раньше даже не задумывались, потому что на первое место все время ставилось выполнение роботом физической работы.

Посмотрите на робота Baxter от компании Rethink Robotics. Это промышленный робот, предназначенный для совместной с людьми работы на конвейере. У него есть лицо, по выражению которого можно понять, что робот собирается предпринять дальше. Так как оценки и прогнозы в данном случае делаются путем считывания невербальных сигналов, робот использует принятую у людей систему сигнализации.

М. Ф.: Как продвигаются разработки сильного ИИ? И можно ли его создать?

С. Б.: Моделирование человеческого интеллекта — один из способов понять его. Я всегда считала, что машины, которые ведут себя как люди, так же принимают решения и так же воспринимают мир, сделают нашу жизнь лучше. Но должны ли эти роботы быть именно людьми? Я так не думаю. Людей и так много. Поэтому нуж-

но нечто, позволяющее расширять наши возможности. Машины, которые всю работу возьмут на себя, не нужны, потому что люди перестанут развиваться. Необходимость оставить пространство для развития накладывает довольно серьезные ограничения на характер отношений человека с машиной.

М. Ф.: Какие технологические прорывы необходимы для приближения к сильному ИИ?

С. Б.: Мы уже научились создавать ИИ специального назначения, который можно обучить настолько, что в узкой предметной области он намного превзойдет человека по эффективности. Но для универсального ИИ требуются принципиально другие виды интеллекта. Мы пока не знаем, как построить машину, которая может развиваться, как ребенок, и все время умнеть.

Ряд прорывов был совершен благодаря глубокому обучению. Но здесь речь идет только об обучении с учителем, в то время как люди учатся разными способами. Мы проектируем машины, обладающие знаниями и информацией в рамках отдельных предметных областей, но понятия не имеем, как заставить их рассуждать на основе здравого смысла, обладать глубоким эмоциональным интеллектом. Список можно продолжить. Предстоит еще много исследований для понимания нашего собственного разума.

М. Ф.: Какие потенциальные опасности несет ИИ?

С. Б.: Основная опасность в данном случае исходит от людей. Меня не так пугает гипотетическое порабощение человечества сверхинтеллектом, как люди, которые могут злоупотреблять новыми технологиями. Нужно думать о вопросах конфиденциальности и безопасности, потому что это связано с нашей свободой. Непонятно, как быть в ситуации, когда массово генерируются фальшивые новости. Опасность представляет и автономное оружие,

и усиление неравенства. Нужно начать работу над демократизацией ИИ, чтобы в будущем эта технология могла принести пользу всем, а не только избранным.

М. Ф.: Стоит ли нам беспокоиться о проблеме выравнивания, или контроля, суперинтеллекта?

С. Б.: Тут первым делом нужно определить, что мы считаем суперинтеллектом. И почему мы считаем, что у него могут быть такие же, как у нас, мотивации и побуждения? Многие связанные с ИИ страхи обусловлены тем багажом, который эволюционно появился в процессе наших попыток выжить в сложном враждебном мире. Откуда взялась идея, что суперинтеллект будет обременен теми же вещами? Это же не человек.

Сама идея о том, что огромное количество талантливых людей будет тратить ресурсы на создание чего-то подобного, абсурдна.

М. Ф.: Но, к примеру, Демис Хассабис из DeepMind крайне заинтересован в создании сильного ИИ. Более того, это официальная цель его компании.

С. Б.: Отдельные люди могут быть заинтересованы в его создании, но откуда взять ресурсы, время и талант в массовом масштабе? Я не понимаю, какие движущие силы в настоящее время могут обеспечить необходимые инвестиции. И не вижу мотивации, которая заставит организации финансировать разработки суперинтеллекта.

М. Ф.: Одним из факторов может быть гонка вооружений.

С. Б.: Соревнование в сфере технологий между разными странами было всегда. Но национальная безопасность не требует сильного ИИ, там вполне можно обойтись более гибким и при этом более специализированным ИИ.

Так что я настаиваю на том, что сейчас гораздо больше причин и мотивации работать над узкими ИИ-приложениями. Конечно, в академических кругах существуют разные исследовательские интересы, и работа над сильным ИИ будет продолжаться. Но практического смысла в нем я пока не вижу.

М. Ф.: Как происходящее повлияет на рынок труда? Ждет ли нас новая промышленная революция и массовая безработица?

С. Б.: Разумеется, такой мощный инструмент, как ИИ, может ускорить технологические изменения. Но пока немногие организации имеют опыт и ресурсы для его развертывания. К сожалению, сейчас нарастает социально-экономический разрыв. И большой вопрос, для чего будет применяться ИИ — для устранения этого разрыва или усугубления. Если только небольшой процент людей знает, как разработать ИИ и как его использовать, получается, что остальные не смогут по-настоящему извлечь выгоду из этой технологии.

Одним из решений этой проблемы является образование. Я прилагаю значительные усилия, внедряя уроки ИИ в школах. Современным детям в будущем предстоит взаимодействовать с интеллектуальными машинами, поэтому крайне важно, чтобы эти машины не воспринимались ими как «черный ящик». В отрасли уже сейчас наблюдается нехватка высококвалифицированных специалистов. Кроме того, люди боятся вещей, которых не понимают. А страх дает возможность для манипуляций.

М. Ф.: Большинство людей не имеет высшего образования, причем ИИ повлияет на них. Должны ли политики вмешаться и урегулировать проблему переподготовки?

С. Б.: Да, сейчас много говорят, например, о том, как повлияют на рынок труда автономные транспортные средства. Фактически человек, рабочее место которого исчезло или видоизменилось, должен пройти переобучение, чтобы остаться конкурентоспособ-

ным. В этом нам может помочь ИИ. Если сделать доступ к приобретению новых навыков и знаний более масштабируемым и гибким, больше людей смогут адаптироваться к меняющимся условиям.

М. Ф.: Нужно ли регулировать сферу ИИ?

С. Б.: В той области исследований, которой занимаюсь я, об этом еще рано говорить. Нужно лучше понять, что такое социальные роботы, прежде чем вырабатывать для них какие-либо правила. Но диалог, который ведется сейчас вокруг ИИ, очень важен, потому что мы начинаем задумываться о последствиях внедрения новых технологий. Я думаю, что начать следует с вещей, которые оказывают особенно сильное влияние на общество, а затем, уже на основе этого опыта думать о дальнейшем регулировании. Очевидно, что нужно искать баланс между сохранением человеческих ценностей и гражданских прав и поддержкой инноваций. Остается определить конкретные детали того пути, который позволит достичь обеих целей.

> *Если бы мы смогли создать оборудование с интеллектом хотя бы на уровне полуторагодовалого ребенка, это стало бы огромным технологическим прорывом.*

Джошуа Тененбаум

ПРОФЕССОР ВЫЧИСЛИТЕЛЬНОЙ КОГНИТИВИСТИКИ В MIT, НАУЧНЫЙ СОТРУДНИК ОБЩЕСТВА ЭКСПЕРИМЕНТАЛЬНЫХ ПСИХОЛОГОВ И ОБЩЕСТВА КОГНИТИВНЫХ НАУК

Джошуа Тененбаум описывает свои исследования как попытку «перепроектировать человеческий разум» и ответить на вопрос: «Каким образом наш мозг извлекает столь многое из столь малого?»

Степень бакалавра физики получил в Йельском университете, а докторскую степень — в MIT. После аспирантуры работал доцентом психологии и computer science в Стэнфорде. Он и его ученики опубликовали множество статей в области когнитивистики, машинного обучения и других областях, связанных с ИИ. Их работы получали награды во всех сферах ИИ, включая компьютерное зрение, обучение с подкреплением, принятие решений, робототехнику, неопределенность в ИИ, когнитивное моделирование и обработку информации в нейронных сетях. Имеет медаль Howard Crosby Warren Medal от Общества психологов-экспериментаторов США, награду за вклад в психологическую подготовку на ранних этапах развития от Американской психологической ассоциации и престижную награду Troland Research Award от Национальной академии наук. Член CSAIL и Центра мозга, разума и машин (CBMM).

Мартин Форд: Сразу начну с главного вопроса: вы считаете, сильный ИИ достижимым?

Джошуа Тененбаум: А что конкретно вы подразумеваете под сильным ИИ? Робота-андроида C-3PO или лейтенанта Дэйту?

М. Ф.: Способность ходить и физически манипулировать вещами я считаю для сильного ИИ необязательной, а вот проходить тест Тьюринга без ограничения по времени он должен. Для меня это некая сущность, с которой можно часами вести умные беседы.

Дж. Т.: Я думаю, создать подобную сущность возможно. Сроки ее появления определяет наш выбор.

М. Ф.: И как продвигается работа над сильным ИИ? Что мешает его созданию?

Дж. Т.: Ответ на второй вопрос зависит от того, какая версия сильного ИИ окажется предпочтительной. Лично я понемногу работаю над той самой системой, с которой можно будет разговаривать часами. Ее интеллект должен быть сравним с нашим по лингвистическим способностям.

Я считаю, что разработку нужно начинать с этапов, предваряющих появление языка, и надстраивать язык на полученную основу. Схему создания описанной вами формы сильного ИИ можно условно разделить на три этапа, соответствующих трем примерным этапам человеческого когнитивного развития.

Первый этап — начальные полтора года жизни, когда закладывается тот интеллект, которым мы обладали в доречевой период. Именно в это время развивается понимание физического мира и действий других людей. Мы это называем наивной физикой, наивной психологией. Эти термины описывают ситуацию, когда вместо теоретического знания используются непосредственные зрительные впечатления. На втором этапе, в возрасте от полутора

до трех лет, формируются языковые навыки, которые применяются для изучения всего остального на третьем этапе.

На мой взгляд, интереснее и ценнее получить умную беседу, проведя систему через все этапы когнитивного развития, и заодно что-то понять в структуре человеческого интеллекта.

М. Ф.: Когда речь заходит о сильном ИИ, люди часто демонстрируют черно-белое восприятие: все, что не интеллект человеческого уровня — это узкоспециализированный ИИ. А вы утверждаете, что вполне возможен некий средний вариант?

Дж. Т.: Да. В докладах я часто показываю видео, в котором полуторагодовалые дети делают удивительно интеллектуальные вещи. И если бы мы смогли построить робота с интеллектом такого уровня, его можно было бы назвать сильным ИИ, ведь он будет иметь гибкое общее представление о доступном ему мире.

Мир взрослого человека простирается от самой ранней истории человечества до будущего, включая в себя множество разных культур, понятых через язык. Но полуторагодовалый ребенок в своем собственном мире непосредственного пространственного и временного окружения обладает гибким, универсальным и практическим интеллектом. Это основное, что нужно осознать.

Современная робототехника добилась большого прогресса в области аппаратного обеспечения. Придуманы алгоритмы управления, позволяющие роботам ходить. Например, вы слышали о компании Boston Dynamics, основанной Марком Райбертом?

М. Ф.: Да, я смотрел видеоролики, на которых их роботы ходят, открывают двери и т. д.

Дж. Т.: И все это сделано по образу и подобию биологических объектов. Марк Райберт всегда хотел понять принцип движения

ног у животных и людей и моделировал движущиеся системы. Его роботы управляются человеком с помощью джойстика. Цели им задает человеческий разум. Интеллект на уровне полуторагодовалого ребенка сделал бы таких роботов огромным технологическим прорывом.

М. Ф.: Какие разработки, направленные на создание сильного ИИ, сейчас самые передовые? Можно ли считать лидером в этой области компанию DeepMind?

Дж. Т.: Я с глубоким уважением отношусь к деятельности DeepMind. Они уже сделали множество классных проектов. Но мне кажется, что к построению ИИ человеческого уровня нужно подходить с другой стороны.

DeepMind в основном сосредоточены на обучении систем с нуля, тогда как люди и животные рождаются с некой предустановленной структурой, определяющей развитие. Для меня она важна.

М. Ф.: То есть вы убеждены в синергии между ИИ и нейробиологией. Как возник ваш интерес к этим областям?

Дж. Т.: Мой отец Джей Мартин Тененбаум был одним из первых исследователей ИИ, получивших докторскую степень в этой области. Он работал в Стэнфорде, где Джон Маккарти начал создавать первую ИИ-лабораторию, занимался компьютерным зрением и принял участие в основании AAAI. По сути, еще ребенком я столкнулся с огромным интересом к ИИ, который имел место в 1970–1980-х гг., и даже посещал тематические конференции. Для одной из таких конференций фирма Apple на один вечер выкупила Диснейленд в Южной Калифорнии — настолько важной сфера ИИ была уже в то время.

Отец некоторое время работал директором ИИ-лаборатории Schlumberger в Пало-Альто. Я часто приходил туда и познакомился со многими лидерами в сфере ИИ. А моя мама Бонни Тененбаум

работала учительницей. У нее была докторская степень в области педагогики. Ее интересовали вопросы обучения и детского интеллекта. Она предлагала мне множество задач и головоломок, которые напоминают то, над чем мы сейчас работаем.

Когда пришло время поступать в колледж, я выбрал физику, параллельно посещая занятия по психологии и философии. Кроме того, я интересовался нейронными сетями, которые тогда, в 1989 г., были популярны.

В 1991 г. я прослушал курс по нейронным сетям. Отец познакомил меня с одним из величайших когнитивных психологов всех времен, своим другом и коллегой из Стэнфорда Роджером Шепардом. Сейчас он на пенсии, а в то время вел научные и математические исследования психических процессов. Я устроился к нему на летнюю работу, где программировал некоторые реализации нейронной сети, теорию которой он разрабатывал. Эта теория рассказывала, как люди и другие организмы решают задачу обобщения.

Роджер пытался математически описать, каким образом организм, получивший от некоего стимула положительные или отрицательные последствия, выясняет, какие еще вещи с большой вероятностью дают те же последствия. Как он выходит за рамки конкретного опыта и формирует общие истины? Как смотрит из прошлого в будущее? Байесовский вывод позволил Роджеру очень элегантно сформулировать свою теорию. Для более масштабируемой реализации этой теории ему требовались нейронные сети. С тех пор я по большей части продолжал работу с этими идеями и методами.

Затем я поступил в аспирантуру в MIT, где до сих пор работаю профессором. После защиты диссертации Роджер помог мне перевестись в Стэнфорд, где я пару лет был ассистентом профессора психологии, прежде чем вернулся в MIT, чтобы заниматься когнитивистикой. Как видите, я пришел в ИИ из естествознания,

но рассматриваю человеческий интеллект с математической, вычислительной и инженерной точек зрения.

Свою деятельность я называю «обратным проектированием ума». Потому что пытаюсь, как инженер, воспроизвести работающий мозг и построить его модель с помощью технических средств. Я рассматриваю ум как невероятную машину, возникшую в результате таких процессов, как биологическая и культурная эволюция, обучение и развитие. Как инженер, я пытаюсь понять, для каких задач предназначен наш мозг и как он их решает.

М. Ф.: Насколько для карьеры в области ИИ важны изучение мозга и когнитивистика? Не кажется ли вам, что computer science уделяется слишком большое внимание?

Дж. Т.: Я всегда рассматривал эти вещи как две стороны одной медали. Меня восхищает сама возможность запрограммировать интеллектуальную машину. Моя специальность не биология, а скорее психология или когнитивистика. Я занимаюсь в основном программным обеспечением интеллекта, а не аппаратными средствами мозга, хотя единственный разумный научный подход предполагает взаимосвязь между ними. Отчасти именно это привело меня в MIT, где есть соответствующий факультет. В середине 1980-х гг. его называли факультетом психологии, но там всегда делался упор на биологию.

Если рассмотреть историю отрасли, окажется, что многие, если не большинство, самых лучших, интересных, новых и оригинальных идей в области ИИ сгенерированы людьми, которые пытались понять, как работает человеческий интеллект. Сюда входят и математические основы того, что мы сейчас называем глубоким обучением и обучением с подкреплением, и изобретение математической логики Джоном Булем, и работа Лапласа по теории вероятностей. Из более поздних примеров можно вспомнить интерес к математике познания и к тому, как люди рассуждают в условиях неопределенности, который привел Джуду Перла к работе над

байесовскими сетями для вероятностного вывода и причинного моделирования в ИИ.

М. Ф.: Вы описали свою работу как попытку «обратного проектирования ума». Как выглядит ее методология? Насколько я знаю, вы много работаете с детьми.

Дж. Т.: Меня с самого начала крайне интересовал вопрос, каким образом наш мозг извлекает столь многое из столь малого. Даже если ребенок не сможет повторить действия, которые ему показали, он все равно поймет, что происходит.

Мы знаем, что корреляция и причинность — не одно и то же и что корреляция не всегда подразумевает причинность. Можно измерить две переменные в наборе данных и увидеть, что они коррелируют, но это не значит, что значение первой обусловливает значение второй. Этот факт часто цитируется, чтобы показать, насколько сложно из данных наблюдений вывести причинно-следственный механизм. Но все же люди, и даже дети, делают это. Посмотрите, как быстро ребенок осваивает управление смартфоном.

Еще студентом вместе с Роджером Шепардом я начал искать, какие же механизмы позволяют людям делать обобщения на базе всего одного или нескольких примеров. Сначала мы использовали принципы байесовской статистики, байесовского вывода и байесовских сетей, то есть формулировали работу ментальных моделей причинно-следственной структуры с помощью теории вероятностей. В 1990-х гг. инструменты, разработанные математиками, физиками и статистиками для статистических выводов на базе небольших наборов данных, стали применяться для машинного обучения, что произвело настоящую революцию. Фактически в сфере ИИ начался переход от ранней символической парадигмы к парадигме статистической.

Затем мы стали задумываться над тем, откуда берутся ментальные модели, и стали изучать интеллект младенцев и детей. К кон-

цу 2000-х гг. мы добились большого прогресса, строя на основе байесовских моделей такие аспекты интеллекта, как восприятие, причинно-следственные связи, а также обнаружение сходства, изучение значений слов, планирование, принятие собственных решений и понимание чужих.

Еще 10 лет назад было построено множество удовлетворительных моделей индивидуальных когнитивных способностей, но объединяющей их теории так и не появилось. Нет у нас и модели здравого смысла.

Если посмотреть, как технологии научились делать вещи, ранее доступные только людям, то можно сказать, что у нас есть настоящий ИИ, просто не такой, каким его задумывали основатели отрасли.

М. Ф.: Это основная цель ваших исследований?

Дж. Т.: Да, в последние годы я действительно заинтересовался универсальным ИИ. И пытаюсь понять, как подобное реализуется с инженерной точки зрения. Сильное влияние на меня оказали исследования моих коллег из Гарварда Сьюзан Кэри и Элизабет Спелке. Они изучали интеллект младенцев и маленьких детей. Я уверен, что на ранней стадии развития работают самые глубокие формы обучения.

С работами Элизабет Спелке должен познакомиться любой, кто собирается заниматься ИИ уровня человека. Она убедительно показала, что уже в возрасте от двух до трех месяцев дети понимают определенные базовые вещи, например, что мир состоит из трехмерных физических объектов, которые не могут просто взять и исчезнуть. Мы называем это свойство постоянством объекта. Раньше считалось, что дети осознают это примерно к году, но Спелке и другие показали, что во многих отношениях наш мозг с самого рождения уже подготовлен к пониманию мира с точки зрения физических объектов и интенциональных агентов.

М. Ф.: Вопрос важности предустановленных структур в ИИ породил множество дискуссий. Показывают ли исследования Спелке, что такие структуры нужны и важны?

Дж. Т.: Идея создания машины, которая изначально обладает интеллектом ребенка и постепенно обучается, была высказана Аланом Тьюрингом в той же статье, где он описал свой тест. Я допускаю, что это одна из самых старых идей в сфере ИИ. Еще в 1950 г. Тьюринг предположил, как построить машину, умеющую проходить его тест, — не стремиться повторить мозг взрослого человека, а начать с детского мозга и постепенно научить его всему. Он сравнил детский мозг с только что купленным блокнотом: маленький механизм и множество чистых листов.

Работы Элизабет Спелке, Рене Байаржон, Лоры Шульц, Элисон Гопник и Сьюзан Кэри показали, что врожденные механизмы обучения невероятно сложны. Сложнее обучения без учителя: дети учатся на гораздо меньшем количестве данных и глубже понимают изучаемые явления.

Меня вдохновляют разработки когнитивистов и психологов о том, как мы воображаем вещи, которых никогда не видели, строим планы и осуществляем их, решая всё новые задачи. Эти разработки показывают, каким образом при обучении применяются ментальные модели: строятся, уточняются и отлаживаются. Наш интеллект годится не только на обнаружение шаблонов в больших данных, он позволяет намного больше.

М. Ф.: Это то, чем вы в последнее время занимались, работая с детьми?

Дж. Т.: Да. Я не считаю свой способ единственно подходящим для построения ИИ-системы, подобной человеку. Но на данный момент это единственный известный нам работающий способ, заодно решающий ряд величайших научных вопросов всех времен о нашей идентичности и понятии «человек».

М. Ф.: Нейронные сети изменили сферу ИИ, но в последнее время часто критикуется ажиотаж вокруг глубокого обучения. Высказываются даже предположения о возможности новой «зимы ИИ». Рассматриваете ли вы глубокое обучение как основной инструмент для продвижения вперед?

Дж. Т.: То, что большинство людей считает глубоким обучением, это один из инструментов в целом наборе. Многие специалисты по глубокому обучению это прекрасно понимают. Но этот термин уже вышел за рамки своего первоначального определения.

М. Ф.: Вместо технического определения, включающего конкретные алгоритмы, такие как метод обратного распространения или градиентный спуск, я бы охарактеризовал глубокое обучение в широком смысле как любой подход, использующий сложные нейронные сети с большим количеством слоев.

Дж. Т.: Для меня использование нейронных сетей с большим количеством слоев тоже всего лишь один инструмент из большого набора. Этот вид глубокого обучения прекрасно подходит для задач распознавания образов, речи, объектов, а также для решения задач, которые можно свести к распознаванию.

Например, исследователи ИИ долгое время не понимали, что задачу воспроизведения игры го можно решить, используя подходы к распознаванию образов. Методы глубокого обучения находят в структуре игрового процесса шаблоны, которые невозможно обнаружить человеку. Но задачу создания сильного ИИ невозможно свести к распознаванию образов.

М. Ф.: Вы пытаетесь преодолеть ограничения?

Дж. Т.: Да, я ищу другие виды инженерных инструментов: графические модели и байесовские сети. Эти вещи имели большое значение, когда я начал работу в сфере ИИ. Одним из самых

важных ученых эпохи «символического ИИ» был Джуда Перл. Тогда, разочаровавшись в символическом интеллекте, вместо статистического подхода все занялись нейронными сетями. Но мне это кажется некорректным, потому что формальные системы подчеркивали силу символических рассуждений и абстрактных языков. К счастью, сейчас начались попытки собрать воедино лучшие идеи из различных парадигм.

Исследования в сфере ИИ можно разделить на три периода — символический, вероятностно-причинный и период нейронных сетей. Это три варианта представлений интеллекта в цифровом виде. У каждого из них были взлеты и падения, просто самый большой успех обеспечили нейронные сети.

М. Ф.: Вы стремитесь к гибриду из нейронных сетей и других, более традиционных подходов?

Дж. Т.: Такие гибриды уже существуют, например, вероятностное программирование. Пока этот инструмент не ассоциируется с ИИ настолько же плотно, как нейронные сети.

В сфере ИИ термины постоянно переопределяются, и хочется сказать, что вероятностные программы примерно так же связаны с вероятностью, как нейронные сети с нейронами. Способы, которыми сейчас используют нейронные сети, выходят далеко за рамки любого нейробиологического процесса. Фактически они давно уже работают с идеями из вероятностных и символических программ. Более того, вероятностные программы приближаются к такому же виду синтеза, только с другой стороны. Потому что для представления здравого смысла нужны не только вероятностные рассуждения, но и абстрактные символические компоненты. Ведь предметные знания — это не просто обмен числами, как в теории вероятностей. Их нужно выражать символами, будь то математика, программирование или логика.

М. Ф.: И вы сконцентрировались на этом подходе?

Дж. Т.: Да, в середине 2000-х гг. мне очень повезло со студентами и аспирантами. Вместе с Ноа Гудманом, Викашем Мансингхкой и Дэном Роем мы создали язык Church, названный в честь математика Алонзо Черча, в котором объединили логические языки высшего порядка, основанные на лямбда-исчислении, что стало основой для универсальных вычислений. Это формальная основа языков программирования, таких как Lisp и его диалект Scheme, которую мы использовали для представления абстрактного знания и обобщения моделей вероятностного и причинного мышления, а затем для построения наивной модели психического состояния человека (способности понять действия других людей на базе их убеждений и желаний). За 10 лет работы с этими вероятностными программами мы впервые смогли построить разумные, количественные, прогнозирующие и концептуально правильные модели понимания чужих действий. Рассмотрели мы и обратный процесс, когда из наблюдений за другими людьми делались выводы об их желаниях и убеждениях. Это пример базовых логических рассуждений, которые присутствуют даже у маленьких детей. И мы получили убедительные результаты.

М. Ф.: Допускают ли вероятностные методы интеграцию в глубокое обучение?

Дж. Т.: Да, в последние годы этот набор инструментов начали применять в нейронных сетях. Для ускорения выхода нужны нейронные сети и другие виды технологий распознавания образов.

Принцип действия нейронных сетей и вероятностных программ становится все более похожим. Разрабатываются новые языки программирования для ИИ, которые объединяют эти подходы. И в итоге уже не нужно выбирать, что лучше использовать, потому что все они — часть единой языковой структуры.

М. Ф.: Но, к примеру, Джеффри Хинтон не воспринимает всерьез идею гибридного подхода. Мне кажется, что сторонники глубокого обучения думают не только об обучении организма на протяжении всей жизни, но и об эволюции. О том, что в какой-то из ранних форм человеческий мозг был намного ближе к состоянию чистого листа. Возможно, это свидетельствует о естественном возникновении необходимых структур?

Дж. Т.: Без сомнения, человеческий интеллект — в значительной степени продукт эволюции, как биологической, так и культурной. Огромная часть нашей эрудиции и средств проявления интеллекта обусловлена информацией, накопленной поколениями. Ребенок, который вырос на необитаемом острове, конечно, может быть таким же умным, но знать он будет намного меньше, чем мы. Наши тела, появившиеся в результате эволюции, представляют собой невероятно сложные структуры с удивительными функциями. Нет оснований думать, что мозг устроен проще.

Нейробиологи не считают, что мозг — это своего рода чистый лист. Он обладает множеством встроенных структур, включающих в себя как основные модели понимания мира, так и алгоритмы обучения, позволяющие их расширить.

М. Ф.: Можно ли достичь более универсального интеллекта, добавив к глубокому обучению модель эволюционного подхода?

Дж. Т.: В DeepMind сказали бы, что они считают эволюцию частью обучения. Мне кажется, что такой подход имеет право на жизнь при условии, что мы учитываем, как на самом деле работает эволюция. Проблема в том, что в поиске оптимальной архитектуры эволюция создавала сложные структуры, в то время как современные алгоритмы глубокого обучения применяют градиентный спуск по сети фиксированной структуры.

М. Ф.: Как вы считаете, связан ли интеллект с сознанием?

Дж. Т.: Это очень сложный вопрос, потому что разные люди понимают под сознанием разные вещи. Философы, когнитивисты и нейробиологи к единому мнению пока не пришли.

М. Ф.: Хорошо, я переформулирую свой вопрос. Может ли машина обладать своего рода внутренним опытом? Является ли такой опыт обязательным условием для сильного ИИ?

Дж. Т.: Для начала рассмотрим два варианта того, что мы подразумеваем под сознанием. Во-первых, это сенсорные явления любого рода, которые в философии называют термином «квалиа». Это ощущение субъективного опыта, с трудом уловимое в любой формальной системе. Например, два человека называют красный цвет красным потому, что испытают при виде него одинаковые ощущения. Теоретически можно создать машину с такими субъективными переживаниями. Но зачем?

Во-вторых, мы определенным образом ощущаем мир и себя в нем. В таком варианте сознание уже можно рассматривать как вещь, необходимую для интеллекта уровня человека.

Если описать состояние мозга в произвольный момент на уровне функций каждого нейрона, это не будет иметь отношения к субъективному восприятию мира. Для людей мир состоит из объектов, а чувства обеспечивают единое понимание происходящего. Не ясно, как связать этот уровень опыта с нейронами, но подобный опыт необходим сильному ИИ. Причем он описывается на уровне объектов и агентов, а не на уровне состояния нейронов.

Ключевая часть самоощущения — это представление о том, что человек — это нечто большее, чем его тело. Это тема наших исследований. Вместе с философом Лори Полом и моим бывшим студентом и коллегой Томером Уллманом мы работаем над статьей, которую назвали *Reverse Engineering the Self* («Обратное проектирование себя»).

М. Ф.: Это тот же принцип, что и в обратном проектировании интеллекта?

Дж. Т.: Да, мы пытаемся понять один простой аспект нашего «я». Я называю его простым, но это только небольшая частичка набора вещей, которые можно называть сознанием. Специалисты в сфере ИИ, особенно те, кто занимается сильным ИИ, говорят, что работают над машинами, которые могут самостоятельно думать или учиться. Но такая машина обязательно должна себя осознавать.

У современных ИИ-систем, будь то беспилотные автомобили или такие программы, как AlphaGo, ничего подобного нет. Эти системы рекламируются как «самообучающиеся». Но без собственной личности они не понимают, что они делают. Человек, играющий в го, осознает процесс игры. Если он решает научиться играть в го, это будет его решение. Человек может попросить научить его, может практиковаться с другими. Он принимает множество решений, руководствуясь тем, как он себя ощущает в разные моменты времени. Если мы хотим получить систему с интеллектом уровня человека, ее нужно научить делать то, что сейчас делают инженеры. Другой вопрос — нужны ли нам такие машины?

М. Ф.: Какие опасности, на ваш взгляд, несет в себе ИИ в краткосрочной и долгосрочной перспективе?

Дж. Т.: Сейчас много говорят о сингулярности и суперинтеллектуальных машинах, которые решат захватить мир, потому что их цели окажутся несовместимыми с существованием людей. Я допускаю, что в далеком будущем может произойти что-то подобное, но лично меня это не беспокоит. Как я говорил выше, мы пока понятия не имеем, как добавить машинам ощущение себя. Соответственно, мысль о том, что они внезапно решат захватить мир, в настоящий момент абсурдна.

Честно говоря, у нас есть более актуальные поводы для беспокойства. Например, опасности, связанные со все более мощными алгоритмами, которые мы разрабатываем, пытаясь приблизиться к сильному ИИ. Они могут быть использованы как для полезных, так и для злонамеренных действий.

Я думаю, что ИИ-сообщество все больше осознает необходимость заниматься рисками, которые грозят нам в краткосрочной перспективе, будь то вопросы конфиденциальности или прав человека. Или то, как повальная автоматизация повлияет на экономику и рынок труда.

На протяжении почти всей истории существования человечества люди находили какой-то источник заработка и тратили первую часть жизни на приобретение профессиональных навыков. А дальше профессия могла кормить их до самой смерти. При необходимости можно было переквалифицироваться, но большинству этого просто не требовалось.

Сейчас же из-за постоянно развивающихся технологий рабочие места чаще меняются, появляются или исчезают. За время жизни человека это может произойти несколько раз. Разумеется, технологические изменения, из-за которых исчезали целые направления деятельности, случались всегда, но раньше этот процесс занимал намного больше времени. Когда-нибудь всем в течение жизни придется неоднократно переобучаться. И думать об этом следует уже сейчас.

М. Ф.: Возможна ли ситуация, при которой большинство окажется не в состоянии угнаться за постоянно меняющимися технологиями? Не лучше ли рассмотреть идею универсального базового дохода?

Дж. Т.: Я не думаю, что мы столкнемся с подобной ситуацией. Люди — устойчивый и гибкий вид. Да, возможно, наши способности к обучению и переподготовке ограничены. Но сейчас мы

понятия не имеем, какие виды деятельности в будущем получат статус работы и начнут использоваться для получения средств к существованию.

Меня сильнее волнует куда более масштабная и неотложная проблема — изменение климата. Исследования в сфере ИИ способствуют ее приближению. Компьютеры все чаще используются для ИИ-приложений, майнинга криптовалют и прочих вещей, увеличивающих потребление энергии. На мой взгляд, те, кто занимается исследованиями ИИ, должны подумать о том, как их деятельность влияет на климат и как внести позитивный вклад в решение этой проблемы. Сейчас на этот аспект принято не обращать внимания.

Возникают проблемы и в сфере прав человека. Например, ИИ-технологии можно применять для шпионажа. Понятно, что исследователи не могут предотвратить или запретить использование технологий в злонамеренных целях, но они могут придумать какие-то контрмеры, например, дать людям возможность выяснять, когда за ними следят. Но эти моральные проблемы обязательно нужно решать.

М. Ф.: Как вы считаете, способно ли государство гарантировать применение ИИ в позитивном ключе с помощью регуляционных мер?

Дж. Т.: На мой взгляд, в Кремниевой долине работают в основном сторонники доктрины о свободе воли, которые считают, что главное — двигать вперед прогресс, а люди как-нибудь сами разберутся, что со всем этим делать. Честно говоря, мне бы хотелось, чтобы правительство и технологическая отрасль перестали враждебно относиться друг к другу и увидели больше общих целей.

Я оптимист и считаю, что представители разных групп могут и должны больше работать вместе и что исследователи ИИ могут стать инициаторами такого рода сотрудничества.

М. Ф.: Прокомментируйте проблему выравнивания, о которой писал Ник Бостром. С одной стороны, нам еще рано говорить о появлении суперинтеллекта, но может получиться так, что для поиска возможностей контроля таких систем потребуется еще больше времени. И именно поэтому решение проблемы выравнивания нужно искать уже сейчас.

Дж. Т.: На мой взгляд, задумываться о таких вещах можно и даже нужно. Просто не в первую очередь. Хотя бы потому, что существуют намного более актуальные экзистенциальные опасности. Проблему выравнивания ценностей трудно решить, потому что мы толком не знаем, что такое ценности. Те, кто занимается вычислительной когнитивистикой, пытаются понять, что же такое ценности, и выполнить их инженерный анализ. Но моральные принципы невозможно описать с инженерной точки зрения.

Мне кажется, что первым делом нужно лучше понять самих себя, и только потом можно будет переходить к технологической стороне. Нужно осознать, каким образом у людей появляется представление об их ценностях и моральных принципах. Этот вопрос актуален в когнитивистике, и по мере роста машинного интеллекта поиск ответа на него станет общеобязательным.

Кроме того, с помощью ИИ-технологий правительства и компании манипулируют людьми. В сравнении с этой проблемой выравнивание ценностей относится к фундаментальным исследовательским вопросам, которые нам вряд ли придется решать в реальности.

М. Ф.: Удастся ли добиться того, чтобы преимущества ИИ перевесили его отрицательные стороны?

Дж. Т.: По натуре я оптимист, поэтому мне сразу хочется ответить «да», но, к сожалению, этого нельзя гарантировать. Любые технологии, будь то ИИ, смартфоны или социальные сети, преобразуют нашу жизнь, меняют способ взаимодействия друг с другом и природу человеческого опыта. Я не уверен, что все это перемены

к лучшему. Трудно с оптимизмом смотреть на семьи, в которых все проводят время со своими телефонами, или на то плохое, что принесли в нашу жизнь социальные сети.

Мы должны осознать и изучить все способы, которыми технологии делают с нами безумные вещи! Они проникают в наш мозг, меняют нашу систему ценностей, систему вознаграждений, систему социального взаимодействия. Я думаю, что нужно более активно исследовать эти явления.

Мне бы хотелось, чтобы сообщество серьезно озаботилось этой проблемой. Я оптимистично думаю, что в долгосрочной перспективе мы создадим ИИ, который, в конечном счете, принесет пользу всем. Но это возможно только при условии, что мы уже сейчас направим наши усилия в нужную сторону.

> *Большинство людей, не задумываясь, ответит на вопрос «пройдет ли слон в дверной проем?», машина же впадет в ступор. То, что легко для одного, трудно для другого, и наоборот. Я называю это парадоксом ИИ.*

Орен Этциони

ГЕНЕРАЛЬНЫЙ ДИРЕКТОР AI2

Орен Этциони курирует проект Mosaic. Является членом AAAI и успешным предпринимателем, стартапы которого приобретены такими гигантами, как eBay и Microsoft. Участвовал в реализации механизмов поиска, онлайн-шопинга, машинного чтения, извлечения открытой информации и семантического поиска академической литературы. Степень бакалавра computer science получил в Гарварде, а докторскую степень — в Университете Карнеги — Меллона. Был профессором в Университете штата Вашингтон. Стал соавтором более 100 технических документов.

Мартин Форд: Расскажите о проекте Mosaic и других проектах, над которыми вы работаете в AI2.

Орен Этциони: Задача проекта Mosaic — добавить компьютерам здравый смысл. Современные ИИ-системы отлично справляются с узкими задачами. Например, могут обыграть в го абсолютного чемпиона. Но если во время матча начнется пожар, ИИ этого не заметит, потому что не осознает окружающей реальности. В AI2 мы ищем ответ на вопрос, как с помощью ИИ сделать мир лучше. Некоторые сотрудники занимаются фундаментальными исследованиями, но в основном мы решаем инженерные задачи.

Например, в проекте Semantic Scholar рассматривается процесс научного поиска и генерации научных гипотез. Современным ученым приходится читать все больше и больше публикаций, соответственно, они начинают испытывать информационную перегрузку и нуждаются в помощи. Поисковая интернет-платформа Semantic Scholar использует машинное обучение, обработку естественного языка, а также различные методы ИИ, помогая ученым выбирать подходящие статьи.

М. Ф.: Использует ли проект Mosaic символическую логику? В старом проекте Сус люди пытались вручную записать все логические правила, например варианты связи объектов, и система была громоздкой.

О. Э.: Мы собираемся использовать для получения знаний более современные методы — краудсорсинг, обработку естественного языка, машинное обучение и машинное зрение.

Сус реализовывался, если можно так выразиться, задом наперед. Сначала создали репозиторий знаний, а затем на его основе строили логические рассуждения. Мы же первым делом определяем эталон, позволяющий оценить способности программы к рассуждениям с точки зрения здравого смысла. Определение такого эталона нетривиальная задача, но как только мы с ней справимся,

появится возможность эмпирически и экспериментально измерять наш прогресс.

М. Ф.: То есть вы планируете создать своего рода объективный тест на наличие здравого смысла?

О. Э.: Именно так! Если тест Тьюринга проверял, может ли машина мыслить, наш тест будет проверять, насколько здраво.

М. Ф.: Вы также работали над системами, которые пытались сдать экзамен по биологии и по другим предметам. Вы все еще этим занимаетесь?

О. Э.: У Пола Аллена была идея программы, которая читала главу из учебника, а затем отвечала на вопросы. Мы сформулировали аналогичную задачу — написать программу, умеющую проходить стандартизованные тесты. В научном контексте это часть проекта Aristo, а в контексте математических задач — часть проекта Euclid. Мы определили эталонную задачу, а затем начали увеличивать эффективность ее решения.

М. Ф.: Достигли ли вы поставленных целей?

О. Э.: Честно говоря, однозначно на этот вопрос ответить не получится. Мы провели конкурс Kaggle, в котором участвовало несколько тысяч команд со всего мира. Было интересно посмотреть, что мы упустили, но оказалось, что наша технология работала лучше других. То есть результаты можно считать позитивными, но наша программа не получила высокую оценку, ведь для решения задач требовались и зрение, и владение языком. Кроме того, очень мешало отсутствие здравого смысла, что и привело нас к проекту Mosaic.

Тут есть один парадокс. Трудные для людей вещи, например, играть в го на уровне чемпиона мира, не представляют сложности для машин. Большинство людей, не задумываясь, ответит на вопрос «пройдет ли слон в дверной проем?», машина же впадет

в ступор. То, что легко для одного, трудно для другого, и наоборот. Я называю это парадоксом ИИ.

Разработчики стандартизированных тестов хотят взять конкретную концепцию, например фотосинтез, и попросить машину применить ее в определенном контексте. Оказалось, что представить процесс фотосинтеза на уровне ученика шестого класса машине легко. А вот применить эту концепцию к конкретной ситуации она не может, потому что это требует понимания языка и здравого смысла.

М. Ф.: Если получится заложить основы здравого смысла, ускорит ли это прогресс в других областях?

О. Э.: Да. Ведь человек без проблем отвечает на вопрос: «Если растение переместить из темной комнаты ближе к окну, листья начнут расти быстрее, медленнее или же скорость их роста не изменится?», потому что он понимает, что чем больше света, тем быстрее протекает фотосинтез, а значит, рост листьев ускоряется. Компьютеру это неочевидно. Он даже не понимает, что такое «переместить растение ближе к окну».

М. Ф.: Как возник ваш интерес к ИИ и как вы оказались в AI2?

О. Э.: Еще в старшей школе я прочитал книгу «Гедель, Эшер, Бах: эта бесконечная гирлянда»[1], в которой параллельно рассматривалась логика Курта Геделя, живопись Морица Эшера и музыка Иоганна Себастьяна Баха, а также излагались концепции, относящиеся к ИИ, такие как математика и интеллект. С этого и началось мое увлечение. Затем я поступил в Гарвард, в котором тогда стартовал курс по ИИ. Он был не очень глубоким, но в нескольких минутах езды на метро находилась ИИ-лаборатория MIT, в которой преподавал один из ее основателей Марвин Минский. Кроме того, я ходил на семинары Дугласа Хофштадтера.

[1] Хофштадтер Д. Р. Гедель, Эшер, Бах: эта бесконечная гирлянда / Пер. с англ. М. Эскина. — Самара: Бахрах-М, 2001. — 717 с. : ил.

Когда я получил работу на полставки в качестве программиста в ИИ-лаборатории MIT, я был на седьмом небе от счастья. И решил пойти в аспирантуру, чтобы изучать ИИ. Я поступил в Университет Карнеги — Меллона, где работал с одним из отцов-основателей машинного обучения Томом Митчеллом.

Затем я преподавал в Университете штата Вашингтон и участвовал во многих стартапах, связанных с ИИ, так как меня крайне интересовала эта тема. А в 2013 г. ко мне обратились из команды Пола Аллена, предложив создать Институт ИИ. Что и было сделано в январе 2014 г.

М. Ф.: Вы общаетесь с Полом Алленом. Расскажите о его мотивации и видении Института ИИ.

О. Э.: Мне действительно повезло общаться с Полом на протяжении многих лет. Когда я размышлял над предложением возглавить институт, я прочел книгу Пола «Миллиардер из Кремниевой долины»[1]. Она позволила мне ощутить и мощь его интеллекта, и суть его видения. Я понял, что Пол действует в традициях Медичи. Как научный филантроп, он подписал Клятву дарения, публично отдав большую часть своего богатства на благотворительность. Его всегда завораживали идеи ИИ. С 1970-х гг. он задавался вопросом, как добавить машинам семантику и понимание текста.

Пол до сих пор помогает мне формировать видение института, не только оказывая финансовую поддержку, но и давая советы по поводу выбора проектов и направления деятельности института.

М. Ф.: Пол также основал Институт наук о мозге. Вы с ним сотрудничаете?

О. Э.: Да, этот институт был основан еще в 2003 г. Мы называем себя «AI2» потому, что наш институт основан позже. Так как

[1] Аллен П. Миллиардер из Кремниевой долины: история соучредителя Microsoft / Пер. с англ. А. Андреева. — М.: Альпина Бизнес Букс, 2012. — 298 с. : ил.

стратегия Пола состояла именно в создании серии институтов, мы очень тесно обмениваемся информацией. Но наши методологии отличаются. В Институте наук о мозге изучается физическая структура мозга, в то время как мы применяем классическую ИИ-методологию для создания программного обеспечения.

М. Ф.: То есть вы в AI2 не занимаетесь обратным проектированием мозга, а предпочитаете проектировать архитектуру, опираясь на человеческий интеллект?

О. Э.: Совершенно верно. Вспомните, что, пытаясь выяснить, каким образом происходит полет, люди спроектировали самолеты. А теперь у нас есть большие, тяжелые и вместительные пассажирские авиалайнеры Boeing 747, которые совсем не похожи на птиц. Я допускаю, что, когда ИИ будет реализован, он будет работать совсем не так, как человеческий интеллект.

М. Ф.: Как вы относитесь к столь популярным сейчас глубокому обучению и нейронным сетям?

О. Э.: Глубокое обучение обеспечило нам впечатляющие достижения в области машинного перевода, распознавания речи, объектов и лиц. Оно дает прекрасные результаты, когда у нас много маркированных данных и вычислительных мощностей. Но на мой взгляд, люди, считающие, что глубокое обучение — это дорога к сильному ИИ, преувеличивают. Особенно, когда утверждают, что он вот-вот будет создан. Это прекрасный инструмент из имеющегося у нас набора, но он не поможет в моделировании рассуждений, базовых знаний, здравого смысла и многого другого.

М. Ф.: Некоторые утверждают, что при достаточном количестве данных появится возможность усовершенствовать обучение — особенно обучение без учителя, — и мышление с опорой на здравый смысл возникнет органически.

О. Э.: Понятие «развивающийся интеллект» (emergent intelligence) употреблял еще когнитивист Дуглас Хофштадтер. Сейчас его приме-

няют в разных контекстах, когда речь заходит о сознании и здравом смысле. Но какие бы предположения о будущем ИИ мы ни строили, хотелось бы, чтобы они основывались на конкретных наблюдаемых данных. Пока что мы видим, что глубокое обучение применяется как статистические модели высокой емкости, которые улучшаются путем увеличения количества данных. Они представляют собой матрицы чисел, которые умножаются, суммируются, вычитаются и т. д., что никак не способствует появлению здравого смысла или сознания. По крайней мере, я подобного не видел.

М. Ф.: Какие проекты в сфере ИИ можно считать передовыми разработками? Где можно ожидать следующего впечатляющего прорыва?

О. Э.: Мне кажется, что самые захватывающие вещи сейчас делает компания DeepMind.

Меня восхитили успехи программы AlphaZero, потому что ее отличной производительности удалось добиться без множества маркированных примеров. В то же время настольные игры — это очень ограниченная сфера. Мне хотелось бы увидеть достижения в сфере робототехники, обработки естественного языка и переноса обучения.

Интересны и работы Джеффри Хинтона, который пытается искать другие подходы к глубокому обучению. Захватывающие вещи происходят и у нас в AI2: мы хотим добавить к парадигме глубокого обучения символические подходы.

Есть люди, которые занимаются обучением без подготовки (zero-shot learning), когда задачу нужно решать без обучающих этому материалов. Или однократное обучение (one-shot learning), когда программа способна предпринимать некие действия после просмотра всего одного примера. Такими вещами занимается доцент кафедры psychology and data science Нью-Йоркского университета Бренден Лэйк.

Том Митчелл из Университета Карнеги — Меллона занимается непрерывным обучением (lifelong learning), при котором система не только просматривает набор данных и строит модель, но и продолжает функционировать и учиться.

М. Ф.: Существует еще и такая техника, как обучение по плану (curriculum learning), когда процесс начинается с простых вещей и постепенно переходит к более сложным.

О. Э.: Если посмотреть внимательно, ИИ — это область, изобилующая неправильно употребляемыми терминами, порой чрезмерно грандиозными. Изначально использовался не лучший, на мой взгляд, термин «ИИ». Затем появились «человеческое познание» и «машинное обучение». Это красиво звучит, но на самом деле набор используемых здесь методов часто очень ограничен. Фактически мы просто пытаемся расширить сравнительно небольшой набор статистических методов и добавить туда больше характеристик, относящихся к обучению людей.

М. Ф.: Вы верите в возможность создания сильного ИИ?

О. Э.: Конечно. Я материалист и не думаю, что в нашем мозге есть что-то кроме атомов. Следовательно, процесс мышления представляет собой некую форму вычислений. Я вполне допускаю, что через какое-то время мы поймем, как реализовать это в машине, и я допускаю, что нам не хватает интеллекта, чтобы сделать это. Но интуиция подсказывает мне, что сильный ИИ, скорее всего, будет создан. Я не знаю, когда это произойдет, потому что пока мы даже не в состоянии сформулировать вопросы, на которые нужно искать ответ.

Пабло Пикассо говорил, что компьютеры бесполезны, потому что отвечают на вопросы, а не задают их. Строгое определение вопроса означает, что мы можем выразить его математически или как вычислительную задачу. Но есть множество вопросов, для которых строгое определение пока невозможно. Как строго

сформулировать, например, задачу представления естественного языка или здравого смысла внутри компьютера?

М. Ф.: Какие основные препятствия отделяют нас от сильного ИИ?

О. Э.: Для начала нужно научиться создавать программы, умеющие выполнять несколько задач: говорить и видеть, играть в настольные игры, переходить улицу, жевать жвачку… Это, конечно, шуточный пример.

Еще нужны системы, которые учатся на единственном примере и узнают объект по частичной информации.

Также важно самовоспроизведение. Представьте ИИ-систему, имеющую физическое воплощение, которая может быстро делать свои копии. У людей процесс создания себе подобных довольно сложный, но у ИИ-систем он вообще отсутствует. Скопировать программное обеспечение несложно, чего нельзя сказать об аппаратной части.

М. Ф.: Наверное, ключевым можно считать умение переносить знания из одной предметной области в другую. Вы говорили о машине, которая читает главу из учебника, а затем может ответить на вопросы. Мне кажется, именно эта способность лежит в основе истинного интеллекта.

О. Э.: Я с вами полностью согласен. Это тоже шаг на пути к сильному ИИ. Подобная способность нужна для функционирования ИИ в реальном мире, полном непредвиденных ситуаций.

М. Ф.: Я хочу поговорить об опасностях, связанных с ИИ, но сначала скажите, где ИИ может обеспечить самые большие преимущества?

О. Э.: Во-первых, это беспилотные автомобили. Только на американских автомагистралях ежегодно погибает более 35 тысяч человек. За год происходит порядка миллиона несчастных случаев.

Исследования показали, что эти цифры можно уменьшить введением беспилотного транспорта. Меня вдохновляет ИИ, спасающий жизни. Во-вторых, это наука. Ведь именно научные разработки помогают экономическому росту, совершенствованию медицины и другим полезным вещам. Перед нами стоят проблемы рака и устойчивых к антибиотикам вирусов. Проект Semantic Scholar поможет ученым в разработке новых лекарств.

Людям, которые говорят о том, что ИИ лишит нас жизни, мой коллега Эрик Хорвиц отвечает, что отсутствие ИИ-технологий уже сейчас убивает людей. Третья по значимости причина смерти в американских больницах — врачебная ошибка. Многие из этих ошибок можно предотвратить с помощью ИИ. Таким образом, наша неспособность использовать ИИ многим стоит жизни.

М. Ф.: Раз уж вы упомянули беспилотные автомобили, скажите, через какое время, по вашим оценкам, могут появиться такие такси? Когда эта услуга станет общедоступной?

О. Э.: Боюсь, что подобное появится не раньше чем через 10 лет. А может, и через двадцать.

М. Ф.: И в результате множество водителей останутся без работы. И не только они. Вполне возможен экономический спад и исчезновение большого количества рабочих мест.

О. Э.: Я не очень понимаю людей, много и активно говорящих об опасности, которую несет нам суперинтеллект. У нас достаточно реальных проблем. И одна из самых острых — будущее рынка труда. Существует тенденция к сокращению производственных рабочих мест, а внедрение ИИ-технологий ускорит этот процесс.

Но я хотел бы отметить, что в этом аспекте демографическая ситуация работает на нас. Сейчас в среднем уменьшается количество детей и при этом растет продолжительность жизни. Общество стареет. В результате, с одной стороны, мы имеем рост автоматизации, а с другой — количество рабочих рук не увеличивается в том же

темпе, что и раньше. Сейчас прекратился прирост количества работающих женщин. Это дает основания надеяться, что в следующие десятилетия спрос на рынке рабочей силы расти не будет. Но это не решает проблему массовой безработицы из-за автоматизации.

М. Ф.: Поможет ли адаптировать общество к экономическим последствиям автоматизации такая вещь, как универсальный базовый доход?

О. Э.: Ситуация, которую мы уже видели в области сельского хозяйства, явно повторяется в сфере производства. Очевидно, что в ближайшие 10–50 лет многие рабочие места либо полностью исчезнут, либо будут радикально преобразованы. Соответствующие обязанности начнут выполняться меньшим количеством людей, зато намного эффективнее.

Людей, занятых в сельском хозяйстве, сейчас намного меньше, чем раньше, а их рабочие обязанности стали намного сложнее. Естественным образом возникает вопрос, чем заниматься всем остальным? Я не знаю точного ответа, но в феврале 2017 г. я написал статью для журнала *Wired*, которая называется *Workers displaced by automation should try a new job: Caregiver* («Работники, уволенные автоматизацией, должны попробовать новую профессию: социальный работник»)[1], где высказался об уязвимости людей без высшего образования в новой экономической ситуации.

Не поможет нам и универсальный базовый доход, особенно если учесть, что у нас пока нет ни всеобщего здравоохранения, ни всеобщей доступности жилья.

М. Ф.: Фактически получается, что поиск решений в этом случае становится проблемой политиков.

О. Э.: Я не уверен, что существует какое-то универсальное решение, но мне кажется, что имеет смысл рассмотреть варианты

[1] https://www.wired.com/story/workers-displaced-by-automation-should-try-a-new-job-caregiver/

работы, которые ориентированы на человека. Можно представить работу по обеспечению эмоциональной поддержки. Это компаньоны для пожилых людей, это сиделки для детей с особыми потребностями. Есть много групп населения, которые предпочтут видеть рядом с собой человека, а не робота.

Если общество начнет выделять ресурсы на такие виды работы, обеспечив достойную зарплату, на эти рабочие места появится множество претендентов. Разумеется, это не панацея, но подумать в этом направлении имеет смысл.

М. Ф.: О чем еще стоит беспокоиться в ближайшие десятилетия?

О. Э.: Серьезной проблемой уже стали вопросы кибербезопасности, а появление ИИ только усугубляет ситуацию. Беспокоит и возможность создания автономного оружия, особенно самостоятельно принимающего решения. Но проблема массовой безработицы все равно остается на первом месте.

М. Ф.: А что вы думаете по поводу экзистенциальной опасности, которую несет с собой сильный ИИ, и по поводу проблемы выравнивания?

О. Э.: Это интересные вопросы для немногочисленных философов и математиков. Обычным же людям нет никакого смысла волноваться о таких вещах, тем более что в настоящий момент мы не можем предпринять никаких реальных действий для предотвращения подобной угрозы.

С моей точки зрения, если суперинтеллект все-таки появится, это будет очень интересно. Будет здорово получить возможность с ним общаться. Работа над пониманием естественного языка в AI2 кажется мне ценным вкладом в безопасность ИИ. Не менее ценным, чем беспокойство по поводу выравнивания целей. В конечном счете, в обоих случаях речь идет всего лишь о технической проблеме, связанной с целевыми функциями и с обучением с подкреплением. На мой взгляд, мы достаточно инвестируем в безопасность ИИ.

Кроме того, обсуждая ИИ, люди часто упускают из виду разницу между интеллектом и автономностью. Зачастую считается, что эти вещи идут рука об руку[1]. Но в системе AlphaGo, которая не начнет игру, пока ее кто-то не запустит, мы видим высокий интеллект и низкую автономность. А вот группа подростков, пьющая алкоголь в субботу вечером, — это пример высокой автономности при низком интеллекте. Я, конечно, шучу. Реальным примером тут будет, скорее, компьютерный вирус, который, не обладая особым интеллектом, может быстро распространяться по компьютерным сетям. Опасна именно автономность.

М. Ф.: Да, беспилотники или роботы, самостоятельно принимающие решение убить человека, вызывают большое беспокойство в ИИ-сообществе.

О. Э.: Как видите, проблема именно в том, что они могут самостоятельно принимать решения о жизни и смерти. С другой стороны, применяя такие вещи с умом, мы можем спасать жизни. Например, сделав воздействие более целенаправленным. Представьте, что можно устранить террористов, никак не затронув заложников. Кроме того, мы сами проектируем все эти системы и, соответственно, можем выбирать степень их автономности.

М. Ф.: Может ли решить проблему автономных ИИ-систем нормативно-правовое регулирование?

О. Э.: Я считаю, что когда речь заходит о мощных технологиях, регулирование неизбежно и уместно. Но мне кажется, что акцент нужно делать на регулировании ИИ-приложений, а не самой технологии. Обратите внимание, что мы пока не можем четко провести границу между программным обеспечением и ИИ. Но, к примеру, после инцидента с такси Uber Национальный совет по безопасности на транспорте поднял вопрос автомобилей с ИИ-системами. Так что введение регулирующих норм — это только вопрос времени.

[1] https://www.wired.com/2014/12/ai-wont-exterminate-us-it-will-empower-us/

> *ИИ — это лучшее, что с нами когда-либо случалось. Этот факт нужно принять. И использовать ИИ, чтобы раскрыть секреты человеческого мозга. Без него мы с этой задачей не справимся.*

Брайан Джонсон

ПРЕДПРИНИМАТЕЛЬ, ОСНОВАТЕЛЬ КОМПАНИЙ KERNEL И OS FUND

Брайан Джонсон инвестирует в научные проекты, направленные на решение насущных проблем человечества. Миссия компании Kernel — значительное повышение качества жизни за счет увеличения продолжительности здорового состояния. Брайан считает, что будущее человечества будет определяться сочетанием человеческого и искусственного интеллектов (HI + AI).

Основал фонд OS для инвестиций в пользу предпринимателей, совершающих открытия в области геномики, синтетической биологии, ИИ, точной автоматизации и разработки новых материалов. Основанная Брайаном в 2007 г. компания Braintree в 2013-м была приобретена компанией PayPal. Написал детскую книгу Code 7.

Мартин Форд: Как появилась идея компании Kernel и какой вы видите ее в долгосрочной перспективе?

Брайан Джонсон: Обычно компании учреждаются для создания некоего продукта. Я же начал с постановки задачи. Хотелось разработать более совершенные инструменты нейронного кодирования, лечения болезней, прояснения механизмов интеллекта и расширения когнитивных способностей. В настоящее время можно сделать снимок мозга с помощью магнитно-резонансной томографии, прикрепить на кожу электроды и снять электроэнцефалограмму, имплантировать человеку электрод для борьбы с болезнью. Но наш мозг по-прежнему практически недоступен. Я инвестировал 100 млн долларов в компанию Kernel, чтобы выяснить, какие инструменты мы можем разработать. Поиски идут уже два года, и мы намеренно нигде не освещаем результаты. Наша команда из 30 человек прилагает все усилия, чтобы приблизить очередной прорыв. К сожалению, подробнее рассказать о нашей работе я пока не могу.

М. Ф.: Я читал, что вы собирались начать с медицинских приложений для помощи страдающим такими заболеваниями, как эпилепсия. Насколько я понял, вы хотели попробовать активную тактику лечения, включающую в себя операцию на головном мозге, а затем использовать полученную информацию для перехода к более щадящим методам. Это действительно так или вы хотите, чтобы в будущем у каждого появился вживленный в мозг чип?

Б. Дж.: Мы рассматриваем разные возможности получения информации о мозге, но в этом деле все решают вопросы прибыльности бизнеса. Создание имплантируемых чипов — один из вариантов, но у нас есть и много других.

М. Ф.: Как возникла идея основать компании Kernel и OS fund?

Б. Дж.: Все началось, когда я вернулся из миссии мормонов в Эквадоре. Мне был 21 год, и в течение двух лет я видел крайнюю

нищету и страдания. Все время, пока я был в миссии, меня мучил единственный вопрос: что я могу сделать, чтобы принести пользу как можно большему количеству людей? Слава и деньги меня не интересовали. Мне не нравился ни один из тех вариантов, которые я в то время мог себе позволить. Поэтому я решил стать предпринимателем, построить бизнес и к 30 годам уйти на пенсию. В 21 год это казалось хорошим планом. Мне повезло, и четырнадцать лет спустя, в 2013 г., мою компанию Braintree приобрел eBay, заплатив 800 млн долларов наличными.

К этому моменту я уже отошел от мормонизма, который в свое время сформировал мои представления о жизни. Нужно было начинать благие дела. Тогда я не понимал, что люди уже обладают всем, что нужно, чтобы выжить самим и справиться с проблемами.

Большинство управленцев, обладающих капиталом, не имеют научных знаний и инвестируют в понятные им вещи, например, в финансовую или транспортную сферу. Я подумал, что если я, не будучи ученым, начну инвестировать в науку и добьюсь успеха, появится модель, которой смогут следовать другие. Поэтому я инвестировал 100 млн долларов в OS fund, и через пять лет мы вошли в десятку наиболее эффективных американских компаний. Мы сделали 28 инвестиций в технологии, меняющие мир.

Затем появилась идея компании Kernel. Я опросил более 200 умных людей, чтобы узнать, чем они занимаются и почему, и понял: двигатель всех наших действий — это мозг. DARPA и Институт мозга Пола Аллена занимаются по большей части конкретными медицинскими приложениями или фундаментальными исследованиями в области нейробиологии. Никто из опрошенных мной не сказал, что мозг — это самая важная вещь из всего, что нас окружает.

Не существует проектов, нацеленных на то, чтобы прочитать нейронный код и расшифровать наше сознание. В компании Kernel я решил сделать для мозга то, что уже было сделано для генома: секвенирование и разработку инструментов заполнения.

Я хочу иметь возможность читать человеческий мозг и делать в него записи по множеству причин. Мне кажется, что людям нужно радикально повышать свой уровень. Сфера ИИ развивается очень быстро. Мы не знаем, какой функцией можно описать ее рост: линией, буквой S или экспонентой, но в перспективе на графике ИИ окажется вверху и справа.

Люди как вид практически не совершенствуются. Мне иногда возражают, что мы более развиты, чем те, кто жил за пять веков до нас, но по большому счету это не так. Да, мы знаем сложные вещи из физики и математики, но в целом мы точно такие же, как тысячи лет назад. С теми же склонностями. Совершающие те же ошибки. Даже если предположить, что мы совершенствуемся, то все равно в чем-то существенно отстаем от ИИ. На графике мы медленно сдвигаемся левее. Вопрос в том, при какой величине этого разрыва мы начнем ощущать себя неловко.

Кроме того, на нас надвигается кризис рынка труда. Ведутся разговоры об универсальном базовом доходе, на который правительства должны будут каким-то образом найти средства. При этом никто и нигде не говорит о радикальном улучшении человека. Такие разговоры нужно начать. Проблема в том, что в своей массе люди не могут представить себе будущее, так как их воображение ограничено знакомыми вещами.

Мне не нравится, что каким-то образом ИИ стал самой большой угрозой, которая должна волновать всех. Я считаю, что самая большая угроза — это люди, способные причинять друг другу вред. ИИ — это лучшее, что с нами когда-либо случалось. Этот факт нужно принять. И использовать ИИ, чтобы раскрыть секреты человеческого мозга. Без него мы с этой задачей не справимся.

М. Ф.: Вопросами, которые ставит перед собой Kernel, занимаются и другие компании. У Илона Маска есть Neuralink, я думаю, что и Facebook, и DARPA тоже работают в этом направлении. Считаете ли вы их своими конкурентами?

Б. Дж.: В DARPA уже довольно давно изучают мозг, как и в Институте мозга Пола Аллена. Но они не идентифицируют мозг как основную точку входа, ключ ко всему. Мы же рассматриваем его именно с этой стороны. И разрабатываем инструменты для чтения и записи нейронного кода. Чтобы уметь читать и переписывать людей.

Я основал компанию Kernel, а менее чем через год Илон Маск и Марк Цукерберг сделали схожие вещи. Компания Маска выясняет, как переписать человека, чтобы он хорошо взаимодействовал с ИИ, а Facebook изучает взаимодействие с пользователями. Я еще не знаю, ждет ли компании Neuralink, Facebook и Kernel успех в ближайшем будущем, но для отрасли они — обнадеживающее явление.

М. Ф.: И как скоро у нас могут появиться доступные устройства или чипы, позволяющие улучшать человеческий интеллект?

Б. Дж.: Все зависит от того, какая процедура будет реализовываться. До имплантируемых чипов нам еще долго, а вот неинвазивные технологии могут появиться относительно быстро. Я думаю, что через 15 лет нейронные интерфейсы станут такими же распространенными, как сейчас смартфоны.

М. Ф.: Это очень смелое предсказание.

Б. Дж.: Когда я говорю о нейронных интерфейсах, я не указываю их тип. Я не утверждаю, что людям начнут вживлять чипы в мозг. Пользователи просто смогут подключаться напрямую к сети.

М. Ф.: А как вы смотрите на возможность загружать в мозг информацию? Ведь мы пока не понимаем, каким образом мозг хранит знания, соответственно, все это кажется научной фантастикой.

Б. Дж.: Пока подобные вещи никто всерьез не обсуждает. Мы продемонстрировали способы совершенствования обучения или

улучшения памяти, но не способность декодировать мысли. И сказать, когда мы изобретем такую технологию, невозможно.

М. Ф.: Я исследовал вопросы роста безработицы и неравенства в сфере труда из-за автоматизации и стал сторонником базового дохода. Вы же видите решение этой проблемы в развитии когнитивных способностей людей. Но мне кажется, что это не выход.

Повышение осведомленности не поможет людям сохранить доход. Кроме того, люди обладают разным уровнем способностей, и технология, улучшающая познание, не даст всеобщего равенства, а только введет новые критерии конкурентоспособности. Ведь ваши технологии получат только те, кто сможет их оплатить.

Б. Дж.: Люди сильно озабочены вопросами неравенства и влияния на их мозг со стороны правительства и других людей. А когда речь заходит о возможности взаимодействия со своим мозгом, они первым делом начинают думать, что при этом может пойти не так. Будто неравенство — это главная наша беда, и к ней опасно прикасаться.

Но на самом деле мы можем исчезнуть из-за своих разрушительных действий или внешних факторов. Поэтому, на мой взгляд, совершенствование человека как вида нельзя рассматривать как предмет элитного потребления. Тут нет вопроса, каковы плюсы и минусы. Я уверен, что, если мы этого не сделаем, нам грозит вымирание. Разумеется, это не означает, что нужно кидаться головой в омут или принять неравенство как должное.

Просто в этом вопросе следует исходить из принципа абсолютной необходимости. Учитывать интересы каждого члена общества нужно в процессе реализации технологии. Нужно понимать, как построить систему, застрахованную от злоупотреблений. Говорят, что интернет разрабатывался с учетом того, что далеко не

у всех пользователей будут добрые намерения. Сейчас эти вещи не обсуждаются. Обсуждение прекращается на том, что совершенствование считают роскошью. Но это одна из причин наших проблем как вида.

М. Ф.: Фактически вы считаете, что нам придется принять идею еще более радикального неравенства. Потому что сначала потребуется группа людей, которые займутся стоящими перед человечеством проблемами, а уже после того, как эти проблемы будут решены, можно задуматься о том, как сделать, чтобы система работала для всех. Я вас правильно понял?

Б. Дж.: Нет. Я предлагаю развивать технологию. Люди должны модернизировать себя, чтобы оставаться востребованными после появления ИИ, а не исчезнуть как вид. У нас уже есть оружие, позволяющее уничтожить человечество, и в течение десятилетий мы были на грани уничтожения.

Представьте, как люди, живущие в 2050 г., говорят: «Невозможно поверить, что в 2017-м считалось нормальным хранить оружие, позволяющее уничтожить всю планету». Мы сейчас застряли в существующей концепции реальности и не можем вырваться из нее, чтобы сформировать будущее, основанное на гармонии, а не на конкуренции, в котором ресурсов хватит на всех и каждый будет заботиться о процветании общества.

Мы же отталкиваемся от того, что люди всегда будут стремиться причинять друг другу вред. Чтобы преодолеть эти ограничения и наши когнитивные предрассудки, нужны улучшения. Я за то, чтобы улучшать всех одновременно. Конечно, такой подход усложняет процесс развития технологии, но по-другому тут никак.

М. Ф.: У меня возникает ощущение, что вы думаете о повышении не только интеллекта, но и морали. Считаете ли вы, что технологии сделают нас более этичными?

Б. Дж.: Мне в принципе не очень нравится термин «интеллект» применительно к моей концепции. Его обычно связывают с IQ, что, на мой взгляд, некорректно. Говоря о радикальном улучшении людей, я имею в виду все возможные сферы.

Если ИИ возьмет на себя большую часть повседневных задач, мозг человека освободится от вещей, которым он посвящает 80 % своих ресурсов. И у него появится время на задачи высшего порядка. Что мы тогда будем делать? Вдруг изучение физики и квантовой теории начнет восприниматься как такая же интересная и приятная задача, как просмотр телешоу? Что мы тогда сможем создать? И к чему придем?

Концепция, которую я предлагаю, сложна для понимания, потому что мозг убеждает нас, что мы уже понимаем все вокруг, и текущая реальность является единственно возможной. Гутенберг не мог представить все книги, которые будут изданы благодаря его изобретению.

Людям нужно смириться с ИИ и потихоньку готовить почву для будущих трансформаций.

М. Ф.: Как вы предлагаете бороться с последствиями введения вашей технологии, особенно в условиях демократии? Достаточно посмотреть на социальные сети, с которыми возникло множество проблем. Ведь речь идет о новом уровне социального взаимодействия.

Б. Дж.: Ничего неожиданного не будет. Люди вполне предсказуемо начнут использовать новые инструменты в собственных интересах. Для заработка, получения статуса, уважения и преимущества перед другими. Это было, есть и будет… если мы не займемся улучшением себя. Если не сможем выйти за собственные пределы и сформировать человечество 3.0 или 4.0. Проблема в том, что прямо сейчас у нас нет инструментов для подобных улучшений.

М. Ф.: Но то, что вы предлагаете, невозможно без регулирования со стороны. Потому что, скорее всего, многие не захотят повышать именно моральный уровень, если сочтут его невыгодным.

Б. Дж.: На мой взгляд, к этой проблеме нужно подходить иначе. В вашей формулировке вопроса подразумевается, что правительство — единственная группа, способная разрешить конфликт интересов. На самом деле это не так. Люди сами могут создавать сообщества для регулирования происходящего.

Во-вторых, вы утверждаете, что каждый человек имеет право сам выбирать, какой уровень морали и этики ему нужен. Но если оглянуться назад, мы увидим, что почти все биологические виды, существовавшие в течение четырех с лишним миллиардов лет, уже вымерли. Люди сейчас в трудном положении, и каждый должен это осознать.

Например, недавно вышли книги *Factfulness: Ten reasons we're wrong about the world, and Why things are better than you think* («Фактологичность: десять причин, по которым мы ошибаемся насчет мира, и почему все лучше, чем кажется») Ханса Рослинга и *The Better Angels of Our Nature: Why Violence Has Declined* («Высшие ангелы нашей природы: почему насилие уменьшилось») Стивена Пинкера. В этих книгах утверждается, что по всем данным мир становится лучше, и этот процесс ускоряется. Но они не учитывают, что у людей никогда не было такой быстро развивающейся формы интеллекта, как ИИ. Не было таких разрушительных инструментов. Поэтому я не принимаю исторический детерминизм как доказательство развития человечества. Мы должны меняться.

Люди же продолжают действовать наобум. Зачастую мы обращаем внимание на то, что происходит вокруг, только после того, как наступит кризисная ситуация. Но невозможно двигаться вперед, ничего не планируя. Мы же как вид ничего не планируем. А уже пора задуматься, как сохранить то, что нам важно.

М. Ф.: Давайте поговорим об ИИ в целом. Используют ли его ваши компании?

Б. Дж.: Да, компании, в которые я инвестировал, используют ИИ и машинное обучение для продвижения научных открытий, чем бы они ни занимались — разработкой лекарств, поиском новых белков или созданием физических продуктов.

Недавно Генри Киссинджер написал в журнал *The Atlantic* открытое письмо, где высказался, что его пугает ИИ в программе AlphaGo, который за несколько дней не просто освоил игру, но и изобрел новые ходы. Реальность Киссинджера напоминает настольную игру, ведь он был политиком во время холодной войны, когда разыгрывалась шахматная партия между США и СССР. Я же только обрадовался тому, что ИИ показывает нам вещи, которых мы не видим. Но большинство людей верит страшным прогнозам и спешат ими поделиться.

М. Ф.: Илон Маск и Ник Бостром выражают обеспокоенность сценарием быстрого взлета и проблемой выравнивания. Они боятся, что ИИ может выйти из-под контроля. Но я слышал, что улучшение когнитивных способностей позволит нам лучше контролировать ИИ. Это действительно так?

Б. Дж.: Я благодарен Нику Бострому за то, что он вдумчиво рассуждает об опасностях, которые может нести с собой ИИ. Он инициировал обсуждения и великолепно все сформулировал. Нужно пытаться предвидеть нежелательные последствия и думать, как их не допустить.

А вот Илон Маск своими непродуманными и крайне эмоциональными высказываниями вызвал негативную реакцию и страх у людей, которые не могут разумно и аргументированно рассуждать на тему ИИ. Мне кажется, что люди как вид должны быть скромнее, признавать собственные когнитивные ограничения и искать способы себя улучшить. Тот факт, что пока эта задача не

является приоритетной, показывает, что нам как виду не хватает смирения.

М. Ф.: Как вы относитесь к международной гонке вооружений в сфере ИИ? Может ли эта конкуренция привести к позитивным результатам? Нужно ли нам менять промышленную политику, чтобы не оказаться в числе отстающих?

Б. Дж.: Так устроен современный мир. Конкурируют отдельные люди и целые государства. Каждый преследует собственные интересы. И если мы не поменяем природу человека, ничего не изменится.

Мне хотелось бы, чтобы в будущем путь к успеху выглядел по-другому. Означает ли это жизнь в гармонии вместо постоянной конкуренции? Может быть. Означает ли это, что этика и мораль изменятся до неузнаваемости? Может быть. Но каждый из нас должен почувствовать собственный потенциал и потенциал всего человеческого рода, чтобы изменить правила игры, которые не приведут нас ни к чему хорошему.

М. Ф.: Вы признаете опасность попадания описываемых вами технологий в недобрые руки. Но глобальное решение в данном случае упирается в проблему координации.

Б. Дж.: Я полностью согласен, это требует нашего пристального внимания и заботы. И именно так поведут себя все государства, потому что так было всегда. Осталось расширить наше воображение до такой степени, чтобы мы могли представить мир, построенный на взаимопомощи.

М. Ф.: Можете ли вы назвать себя оптимистом? Считаете ли, что мы как вид сможем решить стоящие перед нами проблемы?

Б. Дж.: Да, я с оптимизмом смотрю на человечество. Разговоры о трудностях нужны для правильной оценки угрожающих нам опасностей. Я не хочу, чтобы люди прятались от проблем. У нас,

как у вида, есть серьезные проблемы, и я думаю, что нужно пересмотреть наш подход к их решению. Именно для этого я основал OS fund.

Чтобы переосмыслить основные принципы нашего существования, нужно сделать собственное улучшение приоритетной задачей. И для этого нам нужен ИИ. Если мы сможем правильно расставить приоритеты нашего улучшения и максимально привлечем к этому процессу ИИ, обеспечив наш совместный прогресс, мы сможем решить все стоящие перед нами проблемы. И я уверен, что мы создадим фантастическую жизнь для всех, которую сейчас даже не можем себе представить.

Когда появится ИИ человеческого уровня? Результаты опроса

В рамках взятых интервью я попросил всех участников высказать предположение о дате, когда вероятность создания сильного ИИ достигнет как минимум 50 %. Вот результаты этого опроса.

Часть опрошенных не хотела делать предположения, называя конкретный год. Многие говорили, что на пути к сильному ИИ существует множество препятствий. Несмотря на мои усилия, пять человек отказались делать предположение. При этом большинство участников опроса предпочли сделать прогноз анонимно.

Как я отметил во введении, начинают и завершают таблицу ответы двух людей, которые выразили желание указать даты: Рэй Курцвейл — 2029 г. и Родни Брукс — 2200-й.

Вот 18 предположений:

2029	Через 11 лет
2036	Через 18 лет
2038	Через 20 лет
2040	Через 22 года
2068 (3)	Через 50 лет
2080	Через 62 года
2088	Через 70 лет
2098 (2)	Через 80 лет

2118 (3)	Через 100 лет
2168 (2)	Через 150 лет
2188	Через 170 лет
2200	Через 182 года

Среднее арифметическое: 2099 г., через 81 год.

Почти все, с кем я говорил, могли много рассказать о пути к сильному ИИ. И даже те, кто отказался называть конкретную дату, сказали, в каком интервале это может произойти. Поэтому чтение интервью позволяет гораздо лучше понять эту увлекательную тему.

Стоит отметить, что дата, вычисленная как среднее арифметическое, то есть 2099 г., — это пессимистичный результат. На сайте AI Impacts[1] представлены результаты других опросов.

Эти результаты по большей части дают прогноз появления ИИ уровня человека с вероятностью 50 % в диапазоне от 2040 до 2050 г. В большинстве этих опросов участвовало гораздо большее количество человек, а в некоторых случаях опрашивались люди, не имеющие отношения к сфере исследований ИИ.

При этом в небольшой группе, состоящей из высококлассных специалистов, есть оптимисты.

[1] https://aiimpacts.org/ai-timeline-surveys/

Благодарности

Эта книга появилась благодаря командной работе. Редактор издательства Packt Бен Рено-Кларк предложил идею этого проекта в конце 2017 г., и я сразу понял, насколько интересно зафиксировать в одной книге мысли людей, разрабатывающих технологию, которая, скорее всего, изменит наш мир. Бен много сделал для организации проекта и даже участвовал в редактировании отдельных интервью. Моя роль сводилась в основном к организации и проведению бесед. Транскрибацией, а также редактированием и структурированием текста занималась команда издательства Packt, в которую кроме Бена входят Доминик Шекшафт, Алекс Соррентино, Радика Атиткар, Сандип Тадж, Амит Рамадас, Раджвир Самра и Клэр Бойер, которая работала над обложкой.

Я очень благодарен всем опрошенным. Спасибо за время, которые вы потратили на наши беседы, несмотря на свои напряженные рабочие графики. Я надеюсь, что наш проект вдохновит будущих исследователей ИИ и предпринимателей, а также внесет свой вклад в многочисленные обсуждения ИИ, его влияния на общество и нашего потенциала в будущем. Наконец, я благодарю свою жену Сяо-Сяо Чжао и дочь Элейн за их терпение и поддержку во время работы над этим проектом.

Мартин Форд

Архитекторы интеллекта: вся правда об искусственном интеллекте от его создателей

Перевела с английского *И. Рузмайкина*

Заведующая редакцией	*Ю. Сергиенко*
Руководитель проекта	*Н. Римицан*
Ведущий редактор	*К. Тульцева*
Литературный редактор	*А. Руденко*
Художественный редактор	*А. Михеева*
Корректоры	*С. Беляева, М. Молчанова*

Изготовлено в России. Изготовитель: ООО «Прогресс книга».
Место нахождения и фактический адрес: 194044, Россия, г. Санкт-Петербург,
Б. Сампсониевский пр., д. 29А, пом. 52. Тел.: +78127037373.

Дата изготовления: 10.2019. Наименование: книжная продукция. Срок годности: не ограничен.

Налоговая льгота — общероссийский классификатор продукции ОК 034-2014, 58.11.12 —
Книги печатные профессиональные, технические и научные.

Импортер в Беларусь: ООО «ПИТЕР М», 220020, РБ, г. Минск, ул. Тимирязева,
д. 121/3, к. 214, тел./факс: 208 80 01.

Подписано в печать 26.09.19. Формат 70×100/16. Бумага офсетная. Усл. п. л. 33,540. Тираж 1500. Заказ 0000.

ИЗДАТЕЛЬСКИЙ ДОМ ПИТЕР®
WWW.PITER.COM

ИЗДАТЕЛЬСКИЙ ДОМ «ПИТЕР» предлагает профессиональную, популярную и детскую развивающую литературу

Заказать книги оптом можно в наших представительствах

РОССИЯ

Санкт-Петербург: м. «Выборгская», Б. Сампсониевский пр., д. 29а
тел./факс: (812) 703-73-83, 703-73-72; e-mail: sales@piter.com

Москва: м. «Электрозаводская», Семеновская наб., д. 2/1, стр. 1, 6 этаж
тел./факс: (495) 234-38-15; e-mail: sales@msk.piter.com

Воронеж: тел.: 8 951 861-72-70; e-mail: hitsenko@piter.com

Екатеринбург: ул. Толедова, д. 43а; тел./факс: (343) 378-98-41, 378-98-42;
e-mail: office@ekat.piter.com; skype: ekat.manager2

Нижний Новгород: тел.: 8 930 712-75-13; e-mail: yashny@yandex.ru; skype: yashny1

Ростов-на-Дону: ул. Ульяновская, д. 26
тел./факс: (863) 269-91-22, 269-91-30; e-mail: piter-ug@rostov.piter.com

Самара: ул. Молодогвардейская, д. 33а, офис 223
тел./факс: (846) 277-89-79, 277-89-66; e-mail: pitvolga@mail.ru,
pitvolga@samara-ttk.ru

БЕЛАРУСЬ

Минск: ул. Розы Люксембург, д. 163; тел./факс: +37 517 208-80-01, 208-81-25;
e-mail: og@minsk.piter.com

Издательский дом «Питер» приглашает к сотрудничеству авторов:
тел./факс: (812) 703-73-72, (495) 234-38-15; e-mail: ivanovaa@piter.com
Подробная информация здесь: http://www.piter.com/page/avtoru

Издательский дом «Питер» приглашает к сотрудничеству зарубежных торговых партнеров или посредников, имеющих выход на зарубежный рынок: *тел./факс: (812) 703-73-73; e-mail: sales@piter.com*

Заказ книг для вузов и библиотек:
тел./факс: (812) 703-73-73, доб. 6243; e-mail: uchebnik@piter.com

---

Заказ книг по почте: *на сайте www.piter.com; тел.: (812) 703-73-74, доб. 6216; e-mail: books@piter.com*

Вопросы по продаже электронных книг: *тел.: (812) 703-73-74, доб. 6217; e-mail: kuznetsov@piter.com*

КНИГА-ПОЧТОЙ

ЗАКАЗАТЬ КНИГИ ИЗДАТЕЛЬСКОГО ДОМА «ПИТЕР» МОЖНО ЛЮБЫМ УДОБНЫМ ДЛЯ ВАС СПОСОБОМ:

- на нашем сайте: **www.piter.com**
- по электронной почте: **books@piter.com**
- по телефону: **(812) 703-73-74**

ВЫ МОЖЕТЕ ВЫБРАТЬ ЛЮБОЙ УДОБНЫЙ ДЛЯ ВАС СПОСОБ ОПЛАТЫ:

- Наложенным платежом с оплатой при получении в ближайшем почтовом отделении.
- С помощью банковской карты. Во время заказа вы будете перенаправлены на защищенный сервер нашего оператора, где сможете ввести свои данные для оплаты.
- Электронными деньгами. Мы принимаем к оплате Яндекс.Деньги, Webmoney и Qiwi-кошелек.
- В любом банке, распечатав квитанцию, которая формируется автоматически после совершения вами заказа.

ВЫ МОЖЕТЕ ВЫБРАТЬ ЛЮБОЙ УДОБНЫЙ ДЛЯ ВАС СПОСОБ ДОСТАВКИ:

- Посылки отправляются через «Почту России». Отработанная система позволяет нам организовывать доставку ваших покупок максимально быстро. Дату отправления вашей покупки и дату доставки вам сообщат по e-mail.
- Вы можете оформить курьерскую доставку своего заказа (более подробную информацию можно получить на нашем сайте www.piter.com).
- Можно оформить доставку заказа через почтоматы,(адреса почтоматов можно узнать на нашем сайте www.piter.com).

ПРИ ОФОРМЛЕНИИ ЗАКАЗА УКАЖИТЕ:

- фамилию, имя, отчество, телефон, e-mail;
- почтовый индекс, регион, район, населенный пункт, улицу, дом, корпус, квартиру;
- название книги, автора, количество заказываемых экземпляров.

БЕСПЛАТНАЯ ДОСТАВКА:

- курьером по Москве и Санкт-Петербургу при заказе на сумму **от 2000 руб.**
- почтой России при предварительной оплате заказа на сумму **от 2000 руб.**

РЕКОМЕНДУЕМ ПРОЧИТАТЬ

ПИТЕР®

Эта книга представляет собой прикладной подход к анализу текста с помощью Python и охватывает алгоритмы машинного обучения применительно к тексту, классификацию для текстового анализа, структурирование и визуализацию.

КУПИТЬ

Маркос Лопез де Прадо делится тем, что обычно скрывают – самыми прибыльными алгоритмами машинного обучения, которые он использовал на протяжении двух десятилетий, чтобы управлять большими пулами средств самых требовательных инвесторов.

КУПИТЬ

Обработка текстов на естественном языке (Natural Language Processing, NLP) – крайне важная задача в области искусственного интеллекта. Успешная реализация делает возможными такие продукты, как Alexa от Amazon и Google Translate. Эта книга поможет вам изучить PyTorch – библиотеку глубокого обучения для языка Python – один из ведущих инструментов для дата-сайентистов и разработчиков ПО, занимающихся NLP.

КУПИТЬ

Глубокое обучение с подкреплением (Reinforcement Learning) – самое популярное и перспективное направление искусственного интеллекта. Практическое изучение RL на Python поможет освоить не только базовые, но и передовые алгоритмы глубокого обучения с подкреплением. Вы начнете с основных принципов обучения с подкреплением, OpenAI Gym и TensorFlow, познакомьтесь с марковскими цепями, методом Монте-Карло и динамическим программированием.

КУПИТЬ

Промокод на скидку 25 % — Книга