



25
26
27
28
29
30
31
32
33

МАРКУС ДЮ СОТОЙ

КОД КРЕАТИВНОСТИ

Как искусственный интеллект
учится писать, рисовать и думать

Маркус дю Сотой

**Код креативности. Как
искусственный интеллект учится
писать, рисовать и думать**

«Азбука-Аттикус»

2019

УДК 004.8
ББК 32.813

дю Сотой М.

Код креативности. Как искусственный интеллект учится писать, рисовать и думать / М. дю Сотой — «Азбука-Аттикус», 2019

ISBN 978-5-389-18965-2

Знаменитый оксфордский профессор и популяризатор науки Маркус дю Сотой исследует природу творчества, освещая наиболее важные аспекты работы алгоритмов и математических правил, которые лежат в их основе. Он задается вопросом, насколько наш эмоциональный отклик на произведения искусства обусловлен реакцией мозга на закономерности и структуры и что именно означает заниматься творчеством в математике, изобразительном искусстве, литературе и музыке. На основе ярких примеров того, как «повернется алгеброй гармония» мировых шедевров, среди которых «Евгений Онегин» Пушкина, «Песнь льда и пламени» Джорджа Р. Р. Мартина и «Гарри Поттер»; музыкальные произведения Вагнера и Шуберта, с одной стороны, и творчество Massive Attack, Beatles, Боба Дилана — с другой, автор выясняет, через какое время машины, возможно, научатся создавать действительно творческие произведения и могут ли они, в свою очередь, стимулировать наше воображение. Результат перед вами: увлекательное и совершенно необычное исследование как искусственного интеллекта, так и сути того, что значит быть человеком. «Искусственный интеллект потрясает самые основы нашего существования, демонстрируя, сколь многое из того, чем занимаются люди, машины могут делать не хуже, а то и лучше их. Но эта книга посвящена не столько будущему с беспилотными машинами и компьютеризованной медициной, сколько вопросу о том, смогут ли алгоритмы состязаться сколько-нибудь существенным образом с могуществом человеческого кода. Способны ли компьютеры к творчеству? Что нужно для творчества? В какой степени эмоциональная реакция на произведения искусства является результатом реакции мозга на определенные рисунки и структуры? Таковы некоторые из тем, которые мы рассмотрим». (Маркус дю Сотой) В формате PDF А4 сохранен издательский макет книги.

УДК 004.8

ББК 32.813

ISBN 978-5-389-18965-2

© дю Сотой М., 2019

© Азбука-Аттикус, 2019

Содержание

| | |
|-----------------------------------|----|
| 1 | 8 |
| 2 | 13 |
| 3 | 19 |
| Конец ознакомительного фрагмента. | 27 |

Маркус дю Сотой

Код креативности. Как искусственный интеллект учится писать, рисовать и думать

Блестящий путеводитель по грядущему миру искусственного интеллекта.

Джанет Уинтерсон

Интересное, доступное рассмотрение того, до какой степени компьютерам присуще творчество.

Choice

В то время как машины обгоняют нас во все большем числе сфер, мы можем утешить себя тем, что хотя бы одна область останется священной и не поддающейся вычислениям: человеческое творчество. Или все-таки не можем?.. В своем увлекательном исследовании природы творчества Маркус дю Сотой ставит под сомнение многие из подобных предположений. Оксфордский математик, который так же искусен в объяснении сложных теорий в книгах, как и в своих выступлениях на телевидении, утверждает, что многое из того, что мы считаем творчеством, основано на синтезе, осуществляемом сверхразумом, а не порождено вспышкой вдохновения.

Financial Times

Алгоритмы часто рассматриваются как до конца не ясные и опасные силы, и это подпитывает наши страхи перед «духом машин».

Но если искусство – это система раннего предупреждения, а художники – непревзойденные мастера в том, чтобы делать видимым невидимое, то Маркус дю Сотой в своем выдающемся обзоре возможностей ИИ и тех ограничений, которыми связан последний, проливает свет не только на работу программистов и творцов, но и на математику хаоса, лежащую в основе искусства.

*Ханс Ульрих Обрист,
директор лондонской галереи «Серпентайн»*

Эта книга – лишь отчасти об искусстве в его связи с искусственным интеллектом. Как ИИ думает и как он занимается математикой – особый предмет дю Сотоя, который он освещает, как обычно, вдумчиво и глубоко.

Sunday Times

Если все переживания, надежды, мечты, видения, желания, проявления любви и ненависти, которые формируют человеческое воображение, сводятся не более чем к «коду», то рано или поздно машина его взломает. Дю Сотой собрал разнородный массив данных, показывающих, как это происходит уже сейчас.

The Times

Обнадеживающие выводы о том, что истинное творчество принадлежит человечеству... Компьютер может превзойти любого из нас в вычислениях, но ему не хватает того фрагмента «человеческого кода», благодаря которому мы способны понять, что идея не просто нова, а значима.

New York Times Book Review

Далеко идущие рассуждения о значении творчества и мышления.

Wall Street Journal

Насыщенный фактами обзор современных приложений искусственного интеллекта в математике и искусстве.

The Guardian

В своей классической статье 1950 года Алан Тьюринг задается вопросом: «Могут ли машины думать?» Вопрос, который ставит дю Сотой, звучит несколько по-иному, но не менее сложен: могут ли машины быть по-настоящему творческими? Интерес, как и для Тьюринга, заключается не столько в том, чтобы дать исчерпывающий ответ, сколько в глубоком анализе самого вопроса.

Prospect

Захватывающе... Красноречиво и познавательно.

Nature

Алгоритмы, которые не только дублируют человеческие навыки, но и учатся на их ошибках, – вот что определяет искусственный интеллект. Но дю Сотой рассматривает возможность достижения другой стадии: машинного творчества, технологии, которая сама по себе способна к инновациям.

Inside Higher Ed

© Marcus du Sautoy, 2019

© Прокофьев Д. А., перевод на русский язык, 2020

© Издание на русском языке, оформление. ООО «Издательская Группа «Азбука-Аттикус», 2020

КоЛибри®

Посвящается Шани с благодарностью за всю ее любовь и поддержку, творчество и интеллект

1

Тест Лавлейс

Произведения искусства создают правила, но правила не создают произведений искусства¹.

Клод Дебюсси

Машина была прекрасна. На стержни, приводившиеся во вращение поворотом рукоятки, были насажены целые башни шестерней с цифрами на зубцах. Семнадцатилетняя Ада Байрон завороченно вращала рукоятку машины Чарльза Бэббиджа, наблюдая, как та производит расчеты, вычисляет квадраты и даже квадратные корни чисел. Байрон всегда увлекалась машинами, и этот интерес поощряли ее учителя, которых с удовольствием нанимала для нее мать.

Несколько лет спустя, когда Ада, став к тому времени женой графа Лавлейса, изучала планы аналитической машины Бэббиджа, ей пришло в голову, что это устройство – не просто счетный прибор. Она начала составлять перечень потенциальных возможностей этой машины. «Аналитическая машина не имеет ничего общего с обычными “счетными устройствами”. Она занимает совершенно особое место и наводит на более интересные по своей природе соображения».

Заметки Ады Лавлейс считаются теперь первой попыткой создания программного кода. Эта зачаточная идея, развитая революционными достижениями таких людей, как Алан Тьюринг, Марвин Мински и Дональд Мичи, породила ту революцию искусственного интеллекта (ИИ), которая шагает сейчас по всему миру. Однако Лавлейс считала, что возможности любой машины безграничны: «Желательно предостеречь против преувеличения возможностей аналитической машины. Аналитическая машина не претендует на то, чтобы создавать что-то *действительно новое*. Машина может *выполнить все то, что мы умеем ей предписать*»². В конечном итоге, считала она, способности машины ограничены: от нее нельзя получить больше, чем в нее вложено.

Долгие годы эта идея оставалась мантрой информатики. Именно она позволяет нам не бояться, что мы создадим и приведем в действие нечто такое, чем мы не сможем управлять. Некоторые считают, что для создания программы, которая дала бы машине искусственный интеллект, необходимо сначала понять интеллект человеческий.

То, что происходит в наших головах, остается тайной, но в последние годы появилась новая точка зрения на компьютерный код: переход от концепции программирования «нисходящего», идущего от общего к частному, к попыткам идти от частного к общему, по «восходящему» принципу, предполагающему, что компьютер должен сам находить свою дорогу. Оказывается, решения загадки интеллекта не требуется. Можно позволить алгоритмам свободно бродить по цифровому ландшафту и учиться – так же, как учится ребенок. Современные коды, созданные методом машинного обучения, делают совершенно неожиданные вещи – например, они замечают не обнаруженные человеком детали на медицинских изображениях или создают хитроумные инвестиционные стратегии для фондового рынка. Как считают программисты нынешнего поколения, это наконец позволит доказать, что Ада Лавлейс ошибалась: от машины можно получить больше, чем вложено в нее программой.

Тем не менее мы по-прежнему считаем, что одна область человеческой деятельности никогда не станет доступна машинам, – речь идет о творчестве. Мы обладаем необыкновен-

¹ Цит. по: Большая книга мудрости. М.: Эксмо, 2015.

² Цит. по: Гутер Р.С., Полунов Ю.Л. Августа Ада Лавлейс и возникновение программирования // Кибернетика и логика. М., 1978. С. 93, 94 (курсив сохранен).

ной способностью воображать, изобретать и создавать произведения искусства, которые возвышают, расширяют и преображают самую сущность человека. Эти произведения порождаются тем, что я называю «человеческим кодом».

Мы считаем, что этот код присущ только человеку, потому что он является отражением самой сути человека. «Реквием» Моцарта позволяет нам задуматься о нашей собственной смертности. Когда мы смотрим постановку «Отелло», это дает нам возможность исследовать свой собственный эмоциональный мир, царство любви и ревности. Нам кажется, что портрет работы Рембрандта показывает гораздо больше, чем просто внешность модели художника. Можно ли представить себе, что машина когда-нибудь сумеет заменить нам Моцарта, Шекспира или Рембрандта или хотя бы состязаться с ними?

С самого начала я должен подчеркнуть, что по большей части ссылаюсь на примеры художественных произведений западной культуры. Именно это изобразительное искусство я знаю, именно на этой музыке я вырос, именно эту литературу я в основном читал и читаю. Было бы интересно узнать, не окажутся ли машины лучше приспособлены для создания произведений искусства в духе других культур, но я подозреваю, что проблемы, о которых тут идет речь, универсальны и не сводятся к межкультурным различиям. Поэтому, хотя я до некоторой степени готов принести извинения за ограниченность своего кругозора, сосредоточенного на западной цивилизации, я полагаю, что все же смогу дать достаточно общую оценку творческих возможностей наших цифровых конкурентов.

Разумеется, творчество человека не сводится лишь к искусству в традиционном смысле этого слова: к творчеству относятся также и молекулярная гастрономия удостоенного мишленовских звезд повара Хестона Блюментала, и футбольные трюки голландского нападающего Йохана Кройфа, и изогнутые здания Захи Хадид, и знаменитый кубик, изобретенный венгерским инженером Эрнё Рубиком. Даже разработку программного кода для компьютерных игр вроде «Майн-крафт» следует считать частью величайших достижений человеческого творчества.

Несколько менее очевидно то обстоятельство, что творчество играет важную роль и в моем собственном мире – мире математики. Один из факторов, побуждающих меня проводить многие часы за письменным столом, вызывая к жизни уравнения и излагая доказательства, – это привлекательность создания чего-то нового. Моим величайшим творческим достижением, к которому я возвращаюсь снова и снова, была пришедшая мне идея нового симметричного объекта. Никто не знал, что такой объект может существовать. Однако после многих лет упорной работы и одного краткого мига раскаленного добела вдохновения я смог набросать в своем желтом блокноте схему этой невиданной ранее формы. Сама эта волнующая дрожь и есть главный привлекательный момент творчества.

Но что мы на самом деле обозначаем этим изменчивым словом? Те, кто пытается точно определить его значение, обычно отталкиваются от трех основных идей: творчество – это процесс создания чего-то нового, неожиданного и ценного.

Как выясняется, создать нечто новое легко. Я могу заставить свой компьютер обчислить бесконечное число предположений о новых симметричных объектах. Труднее создать нечто неожиданное и ценное. В случае моего симметричного произведения то, что я создал, вполне обоснованно удивило меня – как и других математиков. Никто не предполагал существования открытой мною странной новой связи между этим симметричным объектом и областью теории чисел, не имевшей отношения к моей работе. Именно тот факт, что мой объект позволил получить новое понимание отрасли математики, полной нерешенных задач, определил его ценность.

Все мы привыкаем использовать одни и те же мысленные схемы. Нам кажется, что мы знаем, как будет развиваться сюжет, а потом нас неожиданно уводит в совершенно другом направлении. Этот элемент неожиданности привлекает наше внимание. Возможно, именно

поэтому встреча с творческим произведением, будь оно нашим собственным или чьим-то еще, способна так взволновать нас.

Но что именно придает произведению ценность? Идет ли речь попросту о цене? Необходимо ли признание ценности другими? Я могу считать ценными свое стихотворение или свою картину, но маловероятно, чтобы многие другие люди согласились с моей концепцией их ценности. Необычный роман с множеством неожиданных поворотов сюжета может быть сравнительно малоценным. Но новый и удивительный подход к литературе, архитектуре или музыке, постепенно получающий признание у других и изменяющий наше восприятие вещей, как правило, признается ценным. Именно это Кант называет «образцовой оригинальностью»³, оригинальным деянием, которое становится источником вдохновения для других. Долго считалось, что эта форма творчества присуща только человеку.

Однако на некотором уровне все эти выражения творческого начала – результат нейронных и химических процессов. Они и образуют человеческий код, который в течение миллионов лет оттачивала в нашем мозгу эволюция. Если начать подробно разбирать творческие произведения рода человеческого, можно постепенно увидеть, что в сердце творческого процесса лежат некие правила. Может ли быть так, что творчество основано на алгоритмах и правилах в большей степени, чем нам хотелось бы думать?

Цель этой книги – довести новые модели искусственного интеллекта до пределов их возможностей, чтобы узнать, смогут ли они когда-нибудь создать нечто сравнимое с чудесами нашего человеческого кода или даже превосходящее их. Способна ли машина заниматься живописью, сочинять музыку, писать романы? Возможно, она не станет соперником Моцарта, Шекспира или Пикассо, но сможет ли ее творчество сравниться с творчеством наших детей, пишущих рассказы или рисующих сценки? Сможет ли машина научиться творчеству, взаимодействуя с произведениями искусства, трогающими нас, и выясняя, что именно отличает их от всего приземленного и невыразительного? Более того, сможет ли она раздвинуть границы наших собственных творческих возможностей, показав нам возможности, которых мы не замечаем?

«Творчество» – слово изворотливое и в разных обстоятельствах означает множество разных вещей. Я в основном буду говорить о проблеме творчества в искусстве, но это вовсе не означает, что это единственный возможный род творчества. Мои дочери занимаются творчеством, когда строят замки из LEGO. Когда мой сын приводит к победе свою футбольную команду, его называют творческим полузащитником. Мы можем творчески решать задачи повседневной жизни или творчески управлять работой организаций. И, как я покажу, математика – область гораздо более творческая, чем думают многие, и творчество в этой сфере имеет много общего с творчеством в искусстве.

Творческие импульсы составляют важную часть того, что отличает человека от других животных, и тем не менее мы часто позволяем им застаиваться внутри нас, становимся, сами того не сознавая, рабами установившегося распорядка жизни, рутины. Для творчества необходим толчок, который увел бы нас в сторону от наезженной колеи, которой мы следуем изо дня в день, все более углубляя ее. И тут нам может помочь машина: возможно, она сможет дать нам такой толчок, предложить нам новые возможности, увести нас от простого ежедневного повторения одного и того же алгоритма. В конечном счете машины могут помочь человеку стать менее похожим на машину.

Вы можете спросить, почему именно математик предлагает вам заняться этим исследованием. Простой ответ на этот вопрос заключается в том, что в сердце искусственного интеллекта, машинного обучения, алгоритмов и кодов лежит математика. Чтобы понять, как и почему работают алгоритмы, управляющие современной жизнью, необходимо понять те мате-

³ См.: Кант И. Критика способности суждения / Пер. с нем. М И. Левиной // Собр. соч.: В 8 т. М.: Чоро, 1994. Т. 5. С. 159.

математические правила, на которых они основаны. Тот, кто их не понимает, оказывается безвольной игрушкой машин.

Искусственный интеллект потрясает самые основы нашего существования, демонстрируя, сколь многое из того, чем занимаются люди, машины могут делать не хуже, а то и лучше их. Но эта книга посвящена не столько будущему с беспилотными машинами и компьютеризованной медициной, сколько вопросу о том, смогут ли алгоритмы состязаться сколько-нибудь существенным образом с могуществом человеческого кода. Способны ли компьютеры к творчеству? Что нужно для творчества? В какой степени эмоциональная реакция на произведения искусства является результатом реакции мозга на определенные рисунки и структуры? Таковы некоторые из тем, которые мы рассмотрим.

Но речь идет не просто об обсуждении интересного с интеллектуальной точки зрения вопроса. Художественные произведения людей позволяют получить некоторую информацию о сложном и запутанном человеческом коде, который управляет нашим мозгом; как мы увидим, произведения искусства, которые создают компьютеры, также оказываются поразительно полезны в изучении механизмов работы кода компьютерного. Одна из проблем, связанных с кодом, который формируется таким образом – снизу вверх, – состоит в том, что сами программисты часто не понимают, как именно работает этот код. Почему он принимает то или иное решение? Произведения искусства, которые он создает, могут быть мощным инструментом, позволяющим добраться до подсознательных решений такого нового кода. Кроме того, они же могут выявлять присущие процессу создания этого кода ограничения и опасности, которые мы не вполне осознаем.

Кроме того, я хочу заняться этим исследованием и еще по одной, более личной, причине. Я переживаю сейчас глубокий экзистенциальный кризис. Я то и дело задумываюсь о том, смогут ли люди по-прежнему работать математиками в ближайшие десятилетия с учетом бурного развития искусственного интеллекта. В конце концов, математика – это сфера чисел и логики. Разве не к этому лучше всего способны компьютеры?

Отчасти моя защита от компьютеров, стремящихся «проникнуть» на мою кафедру и «занять места за общим столом», опирается на то соображение, что в математике речь идет не только о числах и логике – это в высшей степени творческое занятие, в котором важную роль играют красота и эстетика. В этой книге я хочу показать, что математические достижения, о которых мы докладываем на своих семинарах и пишем в журналах, появляются не просто в результате того, что человек поворачивает некую механическую рукоятку. Хорошему математику важно иметь интуицию и художественное чувство. И уж конечно, эти черты невозможно запрограммировать в машину. Или все же возможно?

Поэтому я как математик внимательно наблюдаю за тем, насколько успешно новые виды искусственного интеллекта внедряются в художественные галереи, концертные залы и издательства всего мира. Великий немецкий математик Карл Вейерштрасс заметил однажды: «... математик, в котором нет ничего от поэта, никогда не будет настоящим математиком»⁴. Как превосходно показывает пример Ады Лавлейс, в математике должна быть не только частица Бэббиджа, но и частица Байрона. Хотя Лавлейс считала, что возможности машин ограничены, она начинала осознавать, что эти механизмы из зубчатых шестерней могут быть способны и к выражению художественных аспектов мышления:

Этот механизм может действовать не только над числами, но и над другими объектами... Предположим, например, что соотношения между высотами звуков в гармонии и музыкальной композиции поддаются такой

⁴ В письме к Софье Ковалевской от 27 августа 1883 г. – *Здесь и далее, если не указано иное, примеч. перев.*

обработке; тогда машина сможет сочинять искусно составленные музыкальные произведения любой сложности или длительности⁵.

Однако она полагала, что творческое начало будет принадлежать программисту, а не машине. Возможно ли в большей степени переложить эту обязанность на программный код? Программисты нынешнего поколения считают, что это возможно.

На заре развития искусственного интеллекта Алан Тьюринг предложил свой знаменитый тест для измерения разумности компьютера⁶. Я хотел бы предложить новый тест – тест Лавлейс. Чтобы пройти тест Лавлейс, алгоритм должен создать оригинальное произведение искусства, причем этот процесс должен быть повторяемым (то есть не быть результатом случайной аппаратной ошибки), а программист должен быть не способен объяснить, как именно алгоритм пришел к этому результату. Именно такое испытание мы предлагаем машинам – проверку на способность создать нечто новое, неожиданное и ценное. Но для признания наличия у машины подлинно творческих способностей требуется еще одно: произведение машины не должно быть простым выражением творческих способностей автора кода или создателя массива данных. Ада Лавлейс считала эту задачу неразрешимой.

⁵ Цит. по: Гутер Р.С., Полунов Ю.Л. Августа Ада Лавлейс и возникновение программирования // Кибернетика и логика. М., 1978. С. 81.

⁶ Речь идет о статье Тьюринга *Computing Machines and Intelligence* («Вычислительные машины и разум»), опубликованной в 1950 г. в журнале *Mind* и перепечатанной в 1956 г. в четвертом томе сборника «Мир математики» (*The World of Mathematics*) под редакцией Джеймса Р. Ньюмана под названием *Can the Machine Think?* («Может ли машина мыслить?»). Русское издание – «Может ли машина мыслить?» – вышло в 1960 г. в переводе Ю.А. Данилова. В стандартной интерпретации тест Тьюринга формулируется следующим образом: человек взаимодействует с одним компьютером и одним человеком. По ответам, которые он получает от них на свои вопросы, он должен определить, с кем он разговаривает: с человеком или компьютерной программой. Если он приходит к неверному выводу или не может отличить компьютер от человека, считается, что программа прошла тест.

2

Сотворение творчества

*Главный враг творчества – здравый смысл.
Пабло Пикассо*

В связи с тем, как высоко в наше время ценится творчество, самые разные писатели и мыслители пытаются сформулировать, что это такое, как его стимулировать и в чем его значение. Именно во время работы в комитете Королевского общества⁷, созванного для оценки вероятного влияния машинного обучения на общество в ближайшие десятилетия, я впервые познакомился с теориями когнитивиста Маргарет Боден. Ее взгляды на творчество показались мне в высшей степени ценными в контексте проблемы оценки творчества машин.

Боден отличается нестандартным мышлением. За десятилетия своей работы она сумела освоить множество разных дисциплин: она философ, психолог, врач, специалист по искусственному интеллекту и когнитивист. В свои восемьдесят с небольшим Маргарет, с развевающимися сверкающе белыми волосами, по-прежнему активно работает и с удовольствием занимается изучением возможностей этих «жестянок», как она называет компьютеры. Для этого она выделила три разных типа творчества.

Творчество *исследовательское* подразумевает рассмотрение уже существующего и изучение его внешних границ, расширение пределов возможного при соблюдении прежних правил. Музыка Баха стала кульминацией того пути, который прошли композиторы эпохи барокко в исследовании тональности, сплетая в единое целое разные голоса. Его прелюдии и фуги раздвигают границы возможного еще до коренной перестройки жанров и наступления эпохи классицизма Моцарта и Бетховена. Ренуар и Писсарро пересмотрели методы изображения природы и окружающего нас мира, но по-настоящему раздвинул границы Клод Моне, снова и снова писавший свои кувшинки, пока цветковые блики не растворились в новую форму абстракции.

Этим типом творчества увлечена математика. Классификация конечных простых групп – настоящий шедевр исследовательского творчества. Начав с простого определения группы симметрий – структуры, определенной четырьмя простыми аксиомами, – математики в течение 150 лет составляли перечень всех элементов групп симметрий, какие только можно вообразить, и кульминацией этого процесса стало открытие «группы-монстра» *M*, которая содержит больше симметрий, чем содержится атомов в составе Земли, но при этом не соответствует никаким шаблонам других групп. Этот вид математического творчества предполагает расширение границ и в то же время соблюдение установленных правил. Занимающегося им математика можно уподобить путешественнику-исследователю, который продвигается все дальше в неизведанные области, но при этом ограничен пределами планеты.

Боден считает, что человеческое творчество на 97 % состоит из исследования. Компьютеры чрезвычайно способны к творчеству такого рода: доведение некой схемы или набора правил до предела – прекрасная задача для вычислительного механизма, способного выполнить гораздо больше расчетов, чем человеческий мозг. Но достаточно ли этого? Когда мы думаем о по-настоящему оригинальном творчестве, мы обычно представляем себе нечто в большей степени совершенно неожиданное.

Творчество второго типа связано с *комбинированием*. Представьте себе художника, который может взять две совершенно разные концепции и попытаться создать из них нечто единое. Правила, действующие в мире одной из них, часто позволяют создать интересное новое обрам-

⁷ Королевское общество (Royal Society) – британский аналог Академии наук, существует с 1660 г.

ление для другой. Комбинирование – чрезвычайно мощный инструмент в царстве творчества математического. Доказательство гипотезы Пуанкаре, которая описывает возможные формы Вселенной, в конце концов было получено благодаря применению очень разных средств для понимания течения по поверхностям. Потребовался творческий гений Григория Перельмана, чтобы осознать, что характер течения жидкости по поверхности неожиданно может помочь в классификации самих возможных поверхностей.

В моих собственных исследованиях средства, относящиеся к теории чисел, применяются для понимания простых чисел и их использования в классификации возможных симметрий. На первый взгляд кажется, что симметрии геометрических объектов не имеют ничего общего с простыми числами. Но применение языка, который помог нам сориентироваться в тайнах простых чисел, и замена простых чисел симметричными объектами дали нам поразительные новые идеи относительно теории симметрии.

Такого рода «перекрестное опыление» принесло огромную пользу и искусству. Филип Гласс взял идеи, о которых он узнал, работая с Рави Шанкаром, и создал на их основе аддитивный процесс, ставший центральным элементом его минималистской музыки. Заха Хадид разработала уникальный стиль искривленных зданий, скомбинировав познания в области архитектуры со своей любовью к чистым формам русского художника Казимира Малевича. В кулинарии склонные к творчеству повара также объединяют кухни разных концов света.

Имеются интересные данные, позволяющие предположить, что творчество такого рода также может идеально подходить к миру искусственного интеллекта. Если взять алгоритм, играющий блюз, и скомбинировать его с музыкой Булеза, можно получить странное гибридное сочетание, которое, возможно, породит новый мир звуков. Разумеется, может случиться и так, что в результате получится отвратительная какофония. Программисту нужно найти два жанра, которые можно алгоритмически объединить интересным образом.

Более таинственная и трудноопределимая третья форма творчества, по Боден, – творчество *преобразующее*. К этой категории относятся те редкие моменты, которые полностью меняют правила игры. Такие качественные скачки известны во всех видах искусства. Вспомним Пикассо и кубизм, Шёнберга и атональную музыку, Джойса и модернизм. Они подобны фазовым переходам, при которых вода внезапно превращается из жидкости в газ. Именно этот образ нашел Гёте, когда пытался описать, как он в течение двух лет бился над созданием «Страданий юного Вертера», пока не произошло совершенно случайное событие, неожиданно ставшее катализатором его работы: «В это же самое мгновение созрел план “Вертера”; составные части целого устремились со всех сторон, чтобы слиться в плотную массу. Так вода в сосуде, уже близкая к точке замерзания, от малейшего сотрясения превращается в крепкий лед»⁸.

Весьма часто в основе таких преобразующих моментов лежит изменение правил игры или отказ от предположений, из которых исходили в своей работе предыдущие поколения. Квадрат числа всегда положителен. Музыка следует сочинять в ладовой системе. Глаза расположены на лице по разные стороны от носа. На первый взгляд кажется, что запрограммировать такой решительный разрыв было бы трудно, однако для творчества этого типа существует общее правило. Нужно начать с отказа от ограничений и посмотреть, что получится. Искусство, акт творчества состоит в том, чтобы выбрать, что именно следует отбросить или какое новое ограничение ввести так, чтобы получить в результате нечто ценное.

Если бы меня попросили назвать преобразующий момент в математике, я бы назвал хорошей кандидатурой на эту роль открытие квадратного корня из минус единицы, случившееся в середине XVI века. Многие математики считали, что этого числа не существует. Его называли мнимым числом (Декарт придумал это уничижительное название, чтобы подчеркнуть,

⁸ Цит. по: Гёте И.В. Из моей жизни: Поэзия и правда / Пер. с нем. Н.С. Ман // Собр. соч.: В 10 т. М.: Художественная литература, 1976. Т. 3.

что ничего подобного, разумеется, быть не может). И тем не менее его введение не противоречило существовавшей на тот момент математике. Оказалось, что исключать это число из рассмотрения было ошибкой. Но сможет ли компьютер создать концепцию квадратного корня из минус единицы, если все данные, которые в него вводят, говорят о том, что чисел, квадрат которых может быть отрицательным, не существует? Иногда подлинное творчество требует выхода за рамки системы и создания новой реальности. Может ли это быть по силам сложному алгоритму?

История возникновения романтизма в музыке во многих отношениях представляет собой целый каталог нарушения правил. В отличие от композиторов-классиков, использовавших переходы между близкими тональностями, выскочки нового поколения вроде Шуберта предпочитали такие модуляции, которые намеренно не соответствовали ожиданиям. Шуман оставлял неразрешенными аккорды, которые Гайдн или Моцарт просто не могли бы не завершить. Шопен, в свою очередь, сочинял насыщенные хроматические последовательности и необычным образом акцентированные пассажи с непривычными сменами темпа, идущие вразрез с ритмическими ожиданиями. Все переходы от одного течения в музыкальном искусстве к другому – от средневековой музыки к барокко, к классицизму, к романтизму, к импрессионизму, к экспрессионизму и так далее – это сплошное нарушение правил. Творческий уровень каждого следующего художественного течения можно оценить только на фоне предыдущего. Почти само собой разумеется, что важным элементом возможности считать что-либо новым является исторический контекст. Творчество – деятельность не абсолютная, а относительная. Мы творим внутри своей культуры и в своей системе координат.

Способен ли компьютер запустить такого рода фазовый переход и перевести нас в новое музыкальное или математическое состояние? Эта задача кажется трудной. Алгоритмы учатся действовать на основе данных, с которыми они взаимодействуют. Не значит ли это, что они навечно обречены производить только одно и то же?

Как сказал однажды Пикассо, «главный враг творчества – здравый смысл». На первый взгляд кажется, что этот принцип резко противоречит самому духу машины. Однако систему можно запрограммировать на иррациональное поведение. Можно создать метаправило, которое будет заставлять ее менять направление работы. Как мы увидим в дальнейшем, машинное обучение очень хорошо приспособлено к такого рода вещам.

Можно ли научить творчеству?

Многие художники охотно мифологизируют свое творчество, утверждая, что их вдохновение порождается некими сторонними силами. В Древней Греции считалось, что поэты одержимы музами, которые внедряют в их разум вдохновение, тем самым иногда доводя их до безумия. Платон считал, что «поэт – это существо легкое, крылатое и священное; и он может творить лишь тогда, когда сделается вдохновенным и исступленным и не будет в нем более рассудка... И вот поэты творят и говорят много прекрасного... не с помощью искусства, а по божественному определению»⁹. Сходным образом великий индийский математик Рамануджан говорил, что его великие открытия основаны на идеях, которые он получил во сне от богини Намагири, покровительствовавшей его семье. Что же такое способность к творчеству – форма безумия или божественный дар?

К числу самых искушенных специалистов по части запутывания следов своего творчества принадлежит один из моих кумиров в истории математики, Карл Фридрих Гаусс. В 1798 году он опубликовал одну из величайших математических работ всех времен, в которой, как принято считать, создал современную теорию чисел, – трактат под названием *Disquisitiones*

⁹ Платон. *Ион* / Пер. с др. – греч. Я.М. Боровского // Собр. соч.: В 4 т. М.: Мысль, 1990. Т. 1.

*arithmeticae*¹⁰. Когда читатели книги Гаусса попытались понять из нее, откуда он взял свои идеи, они совершенно зашли в тупик. Сама работа была представлена в книге как тайна за семью печатями. Казалось, что Гаусс извлекает откуда-то идеи, как кроликов из шляпы, нигде не давая читателю ни малейшего намека на то, как именно он делает свои фокусы. Позднее, отвечая на упреки в излишней скрытности, он ответил, что архитектор не оставляет строительных лесов после завершения постройки. Подобно Рамануджану Гаусс приписывал одно из своих откровений «милости Божией» и говорил: «Будто молния сверкнула – решилась загадка; я сам не мог найти связующей нити между своими прежними знаниями и последними исследованиями и тем способом, каким она была, наконец, решена»¹¹.

Однако тот факт, что художник может быть не способен внятно изложить, откуда взялись его идеи, не означает, что он не следует никаким правилам. Искусство есть сознательное выражение мириады логических элементов, которые формируют процесс нашего бессознательного мышления. Разумеется, мысли Гаусса были соединены некой логической цепочкой: ему просто было трудно сформулировать, что именно он делает, – или же, возможно, он хотел сохранить тайну, пытаясь раздуть свою славу гениального творца. Утверждение Кольриджа о том, что навеянное «приемом болеутоляющего»¹² видение о Кубла Хане явилось ему сразу целиком, опровергают все те подготовительные материалы, которые свидетельствуют о работе, проделанной поэтом до того судьбоносного дня, в который его разбудил неожиданный посетитель из Порлока. Разумеется, так получается более интересная история. Даже в рассказе о моем собственном творчестве вспыхкам вдохновения неизбежно было бы уделено больше внимания, чем многолетней подготовительной работе.

У нас есть ужасная привычка романтизировать гениальных творцов. Честно говоря, образ одинокого художника, творящего в изоляции, – миф. Брайан Ино говорит о концепции не гения, а «сцения»¹³, чтобы подчеркнуть роль общества, в котором часто появляется творческое мышление. С этим согласна и американская писательница Джойс Кэрол Оутс: «Искусство, как наука, следует считать совместным усилием – попыткой одного человека озвучить мысли многих, попыткой синтезировать, изучать и анализировать»¹⁴.

Что же требуется для стимулирования творчества? Может ли существовать возможность запрограммировать его в машину? И существуют ли правила, следуя которым можно приобрести творческие способности? Другими словами, может ли творчество быть приобретенным навыком? Кое-кто скажет, что такое обучение, или программирование, сводится к обучению подражанию достигнутому ранее и что имитация и следование правилам несовместимы с творчеством. Тем не менее у нас есть множество примеров творческих личностей, повышающих свое мастерство за счет исследования и обучения. Можно ли, исследовав их деятельность, научиться повторять их достижения и в конце концов развить собственные творческие способности?

Эти вопросы я задаю себе в начале каждого семестра. Чтобы получить докторскую степень, аспирант-математик должен создать новое математическое построение. Это значит, что он должен придумать нечто такое, что никогда не было сделано до него. Я должен научить

¹⁰ «Арифметические исследования». Книга эта действительно была написана Гауссом в 1798 г., но опубликована в Лейпциге только в 1801-м. Русский перевод этой и других работ Гаусса (выполненный В.Б. Демьяновым с немецкого перевода): Гаусс К.Ф. Труды по теории чисел // Классики науки. М.: Изд-во АН СССР, 1959.

¹¹ Цит. по: Бюлер В. Гаусс. Биографическое исследование / Пер. с англ. А.Л. Тоома под ред. С.Г. Гиндикина. М.: Наука. Гл. ред. физ.-мат. лит., 1989. С. 38.

¹² См. вступление к поэме «Кубла Хан, или Видение во сне»: «Вследствие легкого недомогания ему [автору] прописали болеутоляющее средство, от воздействия которого он уснул в кресле...» // Кольридж Сэмюэль Тэйлор. Стихи. М.: Наука, 1974. С. 77. – *Примеч. ред.*

¹³ *Scenius* – неологизм, образованный от английских слов *scene* (сцена) и *genius* (гений).

¹⁴ Цит. по: Иглмен Д., Брандт Э. Креативный вид: как стремление к творчеству меняет мир / Пер. с англ. Ю. Константиновой. М.: Манн, Иванов и Фербер, 2018.

аспирантов, как это сделать. Разумеется, до некоторой степени они обучались этой работе и раньше. Решение задач, даже если их ответ уже известен, тоже требует индивидуального творчества.

Подобное обучение совершенно необходимо для последующего прыжка в неизвестное. Повторяя за другими путь, пройденный к величайшим достижениям, мы надеемся создать среду, которая будет способствовать развитию наших собственных творческих способностей. Однако то, что такой прыжок произойдет, далеко не гарантировано. Я не могу взять с улицы первого попавшегося человека и вырастить из него творческого математика. Возможно, лет за десять обучения мы сможем добиться этого, но, по-видимому, к математическому творчеству способен не каждый мозг. Видимо, некоторые люди могут заниматься творчеством в одной области, но не в другой, хотя понять, что именно делает обладателя одного мозга чемпионом по шахматам, а другого – лауреатом Нобелевской премии по литературе, трудно.

Маргарет Боден признает, что обладать творческим началом не всегда значит быть Шекспиром или Эйнштейном. Она различает, следуя ее терминологии, «творчество психологическое» и «творчество историческое». Многие из нас творят свои личные произведения, делая нечто новое для самих себя, но далеко не новое в историческом контексте. Эти действия Боден и называет моментами психологического творчества. И только путем многократного создания произведений личного творчества можно надеяться в конце концов создать нечто, что другие признают новым и ценным. Историческое творчество – явление редкое, но оно появляется в результате стимулирования творчества психологического.

Моя методика пробуждения в студентах творческого начала основана на тех трех типах творчества, которые выделила Боден. Вероятно, наиболее очевидный путь – исследование. Сначала понять, как мы пришли к нынешнему состоянию вещей, а затем попытаться раздвинуть границы чуть дальше. Это требует глубокого погружения в то, что мы создали до сих пор. Из этого глубинного понимания может возникнуть нечто такое, чего никогда раньше не было. Часто бывает важно внушить студентам, что акт творения очень часто вовсе не похож на Большой взрыв. Он происходит постепенно. Как писал Ван Гог, «великое не создается порывом, а представляет собой цепь постоянно слагающихся малых дел»¹⁵.

Вторую стратегию Боден, стратегию комбинаторного творчества, я считаю мощным средством стимулирования новых идей. Я часто советую своим студентам ходить на семинары и читать статьи по темам, которые кажутся не связанными с теми задачами, над которыми они работают. Рассуждения, относящиеся к совершенно другой части математической вселенной, могут войти в резонанс с решаемой задачей и привести к возникновению свежих идей. Некоторые из наиболее интересных творческих проектов в современной науке реализуются именно на стыке разных дисциплин. Чем больше мы выходим за пределы своих изолированных участков и делимся с другими своими мыслями и затруднениями, тем более творческой может стать наша работа. Именно здесь получают множество быстрых результатов.

На первый взгляд кажется, что творчество преобразующее трудно использовать в качестве стратегии. Но, повторяюсь, задача заключается в проверке существующего положения вещей путем отказа от некоторых ранее установленных ограничений. Нужно попытаться понять, что случится, если изменить одно из основополагающих правил, которые мы привыкли считать частью самой сути рассматриваемого предмета. Такие моменты опасны, потому что этим можно обрушить всю систему, но именно это обстоятельство подводит меня к одной из самых важных составляющих, стимулирующих творчество, – готовности к неудачам.

Если вы не готовы к неудачам, вы не пойдете на риск, который позволил бы вам открыть и создать нечто радикально новое. Именно поэтому система образования и экономическая среда

¹⁵ В письме к брату Тео 22 октября 1882 г. Цит. по: *Van Gogh V. Письма к брату Тео / Пер. П.В. Мелковой. М.: Азбука-классика, 2017.*

– области, в которых терпеть не могут неудач, – бывают столь неблагоприятны для развития творческих способностей. Неудачи учеников важно приветствовать не меньше, чем их успехи. Разумеется, на неудачах диссертацию не защитишь, но они могут научить чрезвычайно многому. В разговорах со своими студентами я снова и снова повторяю призыв Беккета: «Проигрывай. Проигрывай снова. Проигрывай лучше».

Можно ли воплотить эти стратегии в программном коде? В использовавшемся в прошлом нисходящем подходе к программированию было очень мало надежды на проявление творческого начала в результатах работы программы. То, что выдавали созданные программистами алгоритмы, никогда не бывало слишком удивительным для их авторов. Не оставалось возможностей ни для экспериментов, ни для неудач. Но недавно все это изменилось: алгоритм, построенный на коде, который учится на собственных ошибках, сделал нечто новое, ошарашившее его создателей и оказавшееся невероятно ценным. Этот алгоритм победил в игре, которую, по мнению многих, машина в принципе не могла освоить. Игра эта требует творческого подхода.

Именно известие об этом революционном событии и стало причиной моего недавнего экзистенциального кризиса как математика.

3

На старт, внимание... го!

*Мы все конструируем и конструируем,
но интуиция все равно полезна.*

Пауль Клее

Математику часто сравнивают с игрой в шахматы. Между этими двумя занятиями, несомненно, есть связи, но, когда компьютер Deep Blue обыграл лучшего гроссмейстера, какого человечество смогло выставить против него в 1997 году, это не привело к закрытию математических факультетов. Хотя шахматы – хорошая аналогия формального аспекта построения доказательства, есть еще одна игра, по мнению математиков, гораздо более близкая к творческой и интуитивной стороне занятий математикой. Речь идет о китайской игре го¹⁶.

Я впервые познакомился с го, когда был старшекурсником и приехал на математический факультет Кембриджского университета, чтобы выяснить, смогу ли поступить в аспирантуру в поразительную группу, которая участвовала в завершении классификации конечных простых групп, своего рода «периодической таблицы симметрий». Пока я беседовал о будущем математики с Джоном Конвеем и Саймоном Нортонем, входившими в число архитекторов этого великого проекта, меня все время отвлекали сидевшие за соседним столом студенты, которые яростно припечатывали к большой сетке размером 19 × 19 линий, вырезанной на деревянной доске, черные и белые камни.

В конце концов я спросил Конвея, чем это они занимаются. «Это го – самая древняя игра из тех, в которые играют до сих пор». В отличие от шахмат с их воинственным характером, объяснил он, го – игра территориальная. Игроки поочередно ставят на сетку размером 19 × 19 линий белые и черные шашки – «камни». Если вам удастся окружить своими камнями группировку камней противника, его камни становятся вашими. Побеждает игрок, которому к концу партии удалось захватить большее число камней. Казалось, все довольно просто. Тонкость этой игры, объяснил Конвей, заключается в том, что, пытаясь окружить противника, нужно в то же время не дать ему окружить ваши собственные камни.

«Эта игра чем-то похожа на математику: простые правила порождают сложность и красоту». Именно наблюдая за развитием игры между двумя мастерами этого дела, пившими кофе в столовой, Конвей обнаружил в последней части игры – ее эндшпиле – поведение, свойственное новому типу чисел, которые он назвал «сюрреальными».

Я всегда интересовался играми. В любых дальних странствиях я люблю учиться играм, в которые играют местные жители, и привозить их с собой. Поэтому, когда я вернулся из диких кембриджских краев к себе домой в Оксфорд, я решил купить в местном магазине игрушек набор для игры в го и выяснить, чем эта игра так увлекала тамошних студентов. Начав исследовать ее вместе с одним из моих однокашников по Оксфорду, я понял, насколько тонка эта игра. Было очень трудно найти ясную стратегию, которая позволила бы мне выиграть. По мере того как на доску выкладывались все новые камни, казалось, что игра становится все сложнее – в отличие от шахмат, в которых постепенное удаление фигур с доски приводит к упрощению партии.

По оценке Американской ассоциации го, количество возможных партий, не противоречащих правилам игры в го, исчисляется 300-значным числом. Что касается шахмат, инфор-

¹⁶ Принятое в русском языке название «го» (как и названия *go/Go*, используемые во многих европейских языках) происходит от японского названия игры – *и-го*. По-китайски она называется *вэйци*, буквально «облавные шашки». Последнее название часто встречается в русских переводах китайской литературы.

матик Клод Шеннон рассчитал, что для исчисления возможных партий в них должно хватить 120-значного числа (которое называют теперь числом Шеннона). В обоих случаях речь идет о немалых числах, но они дают представление о диапазоне возможных вариантов.

В детстве я много играл в шахматы. Мне нравилось продумывать логические следствия предложенных ходов. Это занятие было по душе росшему во мне математику. Дерево возможных ходов в шахматах ветвится упорядоченным образом, что позволяет компьютеру и даже человеку анализировать вероятные последствия каждого хода, последовательно продвигаясь по разным ветвям. В случае же го, напротив, кажется, что игра не позволяет логически предсказывать последствия будущего хода. Перемещение по дереву возможностей быстро становится невозможным. Это не значит, что игрок в го не обдумывает логические последствия каждого своего хода, но эти рассуждения, по-видимому, сочетаются с более интуитивным ощущением характера партии.

Человеческий мозг активно стремится выискивать в визуальных изображениях структуры и закономерности, если только они там есть. Игрок в го может, глядя на расположение камней и пользуясь способностью мозга находить такие структуры, выбрать свой следующий ход, исходя именно из них. Компьютерам всегда было трудно работать с визуальной информацией. Это одна из тех крупных проблем, над которыми инженеры бьются десятилетиями. Высокоразвитая способность человеческого мозга воспринимать визуальные структуры оттачивалась на протяжении миллионов лет, так как она была совершенно необходима для нашего выживания. Выживание любого животного отчасти зависит от его способности различать в визуальном беспорядке, которым окружает нас природа, закономерности и образы. Упорядоченная структура в хаосе джунглей, вероятно, указывает нам на присутствие другого животного – и ее важно заметить, потому что это животное может нас съесть (а может быть, мы его). Человеческий код чрезвычайно хорошо умеет считывать образы, интерпретировать их возможное развитие и вырабатывать соответствующую реакцию. Эта способность – одно из самых ценных наших преимуществ, и именно она помогает нам понимать и оценивать по достоинству образы в музыке и изобразительном искусстве.

Оказывается, именно распознаванием образов я занимаюсь в своей математической работе, когда отправляюсь в неисследованные уголки математических джунглей. Я не могу просто полагаться на пошаговый логический анализ местной среды. С ним я далеко не уйду. Он должен сочетаться с интуитивным ощущением того, что может находиться где-то рядом. Эта интуиция развивается за время, посвященное исследованию уже известного пространства. Но часто бывает трудно логически аргументировать, почему мне кажется, что в таком-то направлении лежит территория, интересная для исследования. Математическая гипотеза – это, по определению, утверждение еще не доказанное, но у математика, высказывающего гипотезу, уже есть ощущение, что его математическое утверждение может быть хотя бы до некоторой степени истинным. Пробираясь сквозь заросли и пытаюсь прокладывать новые пути, мы используем как наблюдения, так и интуицию.

Математик, умеющий предложить хорошую гипотезу, часто пользуется большим уважением, чем тот, который соединяет логические точки, чтобы продемонстрировать истинность гипотезы. В игре го выигрышная позиция в некоторых отношениях подобна гипотезе, а партия – последовательности логических ходов, которыми игрок эту гипотезу доказывает. Но различить закономерности в процессе игры чертовски трудно.

Поэтому, хотя шахматы действительно помогают понять некоторые аспекты математики, всегда считалось, что го гораздо ближе по духу к тому, как математики на самом деле занимаются своей наукой. Именно поэтому математики не слишком беспокоились, когда компьютер Deep Blue обыгрывал в шахматы лучших представителей рода человеческого. По-настоящему трудной задачей оставалось освоение игры в го. В течение многих десятилетий утверждалось, что компьютер никогда не сможет научиться играть в го. Как и любой порядочный беспреко-

словный постулат, это утверждение побуждало изобретательных программистов попытаться его опровергнуть. Но каждый раз оказывалось, что даже не слишком опытным игрокам удается победить самые замысловатые алгоритмы. Так что математики чувствовали себя в безопасности под прикрытием, которое давала им игра го. Раз компьютеры не могут играть в го, нечего и говорить о том, чтобы они смогли играть в гораздо более тонкую и древнюю игру, которую мы называем математикой.

Однако проломы в конце концов удалось проделать даже в Великой Китайской стене – и моя защитная стена тоже рассыпалась в прах, причем обрушение ее было зрелищем весьма эффектным.

Необыкновенный игрок

В начале 2016 года было объявлено о появлении программы для игры в го, создатели которой были уверены, что она сможет состязаться на равных с лучшими игроками-людьми. Учитывая фиаско всех предыдущих попыток, игроки в го всего мира отнеслись к этому известию чрезвычайно скептически. Тогда компания, разработавшая программу, бросила им вызов. Она организовала открытое соревнование с огромным денежным призом и предложила, чтобы в нем принял участие один из лучших игроков в го всего мира. На это согласился победитель многочисленных международных турниров кореец Ли Седоль. Матч должен был состоять из пяти партий, и победитель получал приз миллион долларов. Имя противника Ли Седоля – AlphaGo.

Программу AlphaGo создал Демис Хассабис. Он родился в Лондоне в 1976 году; его отец был греком с Кипра, а мать происходила из Сингапура. Оба его родителя были преподавателями и, по словам самого Хассабиса, богемными технофобами. Его сестра и брат связали свою жизнь с художественным творчеством: одна стала композитором, другой выбрал своим занятием литературу. Поэтому Хассабис не вполне понимает, как из него получился ученый зануда. Но одаренность и талантливость Хассабиса были отмечены еще в раннем детстве, особенно там, где дело касалось игр. В шахматах он проявил такие способности, что в одиннадцать лет занимал второе место в мировом рейтинге детей своего возраста.

Однако именно в этом возрасте на международном турнире в Лихтенштейне Хассабиса озарило: чем это все они занимаются? Зал был заполнен множеством первоклассных умов, которые исследовали логические хитросплетения великой игры. Но Хассабис внезапно осознал полную бессмысленность этого занятия. В интервью радиостанции Би-би-си он рассказал, что в тот момент подумал: «Мы растрачиваем свой разум впустую. Почему бы нам не направить всю эту интеллектуальную мощь на что-нибудь более полезное, например на борьбу с раком?»

После этого турнира (в котором он чуть было не победил, уступив голландскому чемпиону мира среди взрослых после десятичасовой борьбы) он шокировал своих родителей, заявив, что прекращает участвовать в шахматных соревнованиях. Все думали, что именно это станет делом всей его жизни. Но годы, посвященные шахматам, не были потрачены впустую. За несколько лет до того он потратил 200 фунтов денежного приза, полученного за победу над противником из США Алексом Чангом, на покупку первого в своей жизни компьютера – ZX Spectrum. Этот компьютер породил в нем страстное стремление добиться того, чтобы за него думали машины.

Вскоре Хассабис перешел на следующий уровень: у него появился Commodore Amiga. Этот компьютер уже позволял программировать игры, которые ему нравились. Шахматы были слишком сложны, но Хассабис сумел научить свой Commodore играть в «Отелло»¹⁷, игру, довольно похожую на го, – в ней используются черно-белые камни, которые переворачивают,

¹⁷ Один из вариантов игры реверси.

когда они оказываются заперты в окружении камней другого цвета. Эта игра слишком проста для гроссмейстеров, и Хассабис испытывал свою программу на собственном младшем брате. Программа неизменно обыгрывала его.

Это было классическое программирование по принципу «если... то...». Нужно было вручную запрограммировать реакцию на каждый возможный ход противника: «если противник делает такой-то ход, мы делаем такой-то ответный ход». Все творческое содержание было обеспечено Хассабисом и его способностью угадывать правильные ответные ходы, необходимые для победы в игре. И все равно казалось, что в этом есть какое-то волшебство. Стоило запрограммировать правильное заклинание, и Commodore, подобно ученику чародея, делал все, что требовалось, чтобы довести игру до победы.

Хассабис стремительно закончил школу и уже в шестнадцать лет получил предложение изучать информатику в Кембридже. Кембридж завоевал его сердце, когда он посмотрел фильм «История жизни»¹⁸ с Джеффом Голдблумом. «Я подумал: в Кембридже происходит вот такое? Можно туда поехать и открыть ДНК, сидя в пабе? Ух ты!»

В шестнадцать лет он еще не мог начать учиться в Кембридже, так что ему пришлось отложить поступление на год. Чтобы занять это время, он, завоевав второе место в конкурсе, который проводил журнал Amiga Power, устроился на работу в компанию, разрабатывавшую компьютерные игры. Там он создал свою собственную игру «Тематический парк» (Theme Park), в которой игроки должны были создать собственный тематический парк и управлять его работой. Игра имела огромный успех: она разошлась миллионными тиражами и была удостоена премии «Золотой джойстик» (Golden Joystick). Накопленных денег должно было хватить на время обучения в университете, и Хассабис отправился в Кембридж.

Учебный курс познакомил его с великими деятелями революции искусственного интеллекта – Аланом Тьюрингом и его тестом на разумность, Артуром Сэмюэлом и его программой для игры в шашки, Джоном Маккарти, который и придумал термин «искусственный интеллект», Фрэнком Розенблаттом и его первыми опытами с нейронными сетями. Именно они были теми гигантами, на плечах которых стремился стоять Хассабис. Именно на лекциях в Кембридже он слышал, как его преподаватель повторяет как мантру, что компьютеры никогда не смогут играть в го из-за творческих и интуитивных аспектов этой игры. Это утверждение подействовало на молодого Хассабиса, как красная тряпка на быка. Он покинул Кембридж, преисполненный решимости доказать, что преподаватель ошибался.

Его идея состояла вот в чем: нужно попытаться написать не саму программу, способную играть в го, а некую метапрограмму, которая, в свою очередь, будет писать программу, играющую в го. Идея эта казалась безумной, но суть ее сводилась к тому, что метапрограмма должна быть создана таким образом, чтобы она могла учиться на своих собственных ошибках по мере того, как играющая программа проводит все больше партий.

Хассабис узнал о реализации похожей идеи в 1960-х годах исследователем искусственного интеллекта Дональдом Мики. Мики написал алгоритм под названием MENACE¹⁹, который, начав с нуля, научился оптимальной стратегии игры в крестики-нолики. Название MENACE было акронимом слов *Matchbox Educable Noughts And Crosses Engine* – «Обучаемое устройство из спичечных коробков для игры в крестики-нолики». Для демонстрации работы своего алгоритма Мики собрал систему из 304 спичечных коробков, представлявших все возможные расположения крестиков и ноликов, встречающиеся в процессе игры. В каждом коробке были разноцветные бусины, представляющие возможные ходы. По окончании каждой партии бусины, соответствующие сделанным ходам, вынимались из коробков в случае проиг-

¹⁸ *Life Story* (в американском прокате – *The Race for the Double Helix*, т. е. «Гонка за двойной спиралью») – телефильм 1987 г. производства Би-би-си, рассказывающий об истории открытия структуры ДНК. Джефф Голдблум сыграл в нем Джеймса Уотсона.

¹⁹ Совпадает со словом *menace* – угроза (*англ.*).

рыша или добавлялись в них в случае победы. По мере увеличения числа партий, сыгранных алгоритмом, распределение бусин по коробкам все больше соответствовало почти совершенной стратегии игры. Именно эту идею обучения на собственных ошибках Хассабис и хотел использовать для тренировки алгоритма в игре в го.

У Хассабиса была хорошая модель, которую можно было положить в основу такой стратегии. Мозг новорожденного младенца не запрограммирован на преодоление всех препятствий, которые встретятся ему в жизни. Вместо этого он запрограммирован на обучение при взаимодействии с окружающей средой.

Если Хассабис собирался использовать в осуществлении своей мечты о создании программы, играющей в го, тот метод, которым мозг обучается решать задачи, ему явно могло помочь знание о том, как работает мозг. Поэтому он решил поступить в аспирантуру по нейробиологии при Университетском колледже Лондона. Именно там, во время перерывов на кофе в процессе лабораторной работы, Хассабис начал обсуждать свои планы создания компании для испытания своих идей с нейробиологом Шейном Леггом. То обстоятельство, что они никогда не рассказывали своим преподавателям о мечте посвятить свою жизнь разработкам искусственного интеллекта, показывает, насколько незавидной была репутация искусственного интеллекта еще лет десять назад. Но они чувствовали, что идут по верному следу, и в сентябре 2010 года двое ученых решили создать вместе с Мустафой Сулейманом, другом детства Хассабиса, свою фирму. Так явилась на свет компания DeepMind.

Компании нужны были деньги, но поначалу Хассабису никак не удавалось привлечь хоть какие-нибудь капиталы. Большинству инвесторов идея компании, посвященной играм и исследованиям интеллекта, казалась недостаточно серьезной. Однако нашлись и такие, кто поверил в эту идею. Среди инвесторов, с самого начала вложивших средства в этот проект, были Илон Маск и Питер Тиль. Тиль никогда не инвестировал за пределами Кремниевой долины и пытался уговорить Хассабиса перебраться на Западное побережье США. Но Хассабис, родившийся и выросший в Лондоне, упорно стоял на своем, утверждая, что в Лондоне гораздо больше невостребованных талантов, которых можно привлечь к делу. Хассабис вспоминает абсурдный разговор, случившийся у него с юристом Тиля. «А в Лондоне есть законы об интеллектуальной собственности?» – невинным тоном спросила она. По-моему, им казалось, что мы из какого-нибудь Тимбукту!» Основателям компании пришлось отдать инвесторам огромную часть ее акций, но зато они получили деньги, позволявшие приступить к решению задачи ИИ.

Создание машины, способной научиться играть в го, все еще казалось отдаленной мечтой. Сперва они взялись за решение задачи, которая представлялась менее заумной: освоение игр Atari 1980-х годов. Вероятно, компания Atari виновата во множестве прогулов уроков учениками конца 1970-х и начала 1980-х. Я лично хорошо помню, как много времени я потратил, играя на приставке Atari 2600 одного своего друга в Pong, Space Invaders и Asteroids. Эта приставка была одним из первых устройств, конструкция которых позволяла играть не в одну, а в несколько разных игр, которые загружались на кассетах. Такая система открыла возможность создать с течением времени целый спектр разнообразных игр. Более ранние приставки позволяли играть только в одну игру, которая была физически встроена в них.

Одна из моих любимых игр на Atari называлась Breakout. Параллельно верхней кромке экрана шла стена из разноцветных кирпичиков, а игрок управлял расположенной внизу ракеткой, которую можно было двигать влево или вправо при помощи джойстика. Шарик отскакивал от ракетки и летел в сторону кирпичиков. Каждый раз, когда он попадал в какой-нибудь кирпичик, кирпичик исчезал. Целью игры было убрать с экрана все кирпичики. За попадание в один из желтых кирпичиков, из которых состояла нижняя часть стены, начислялось одно очко. Красные кирпичики в верхней части приносили по семь очков. По мере исчезновения кирпичиков ракетка становилась все меньше, а шарик летал все быстрее, что делало игру труднее.

Особенное удовольствие мы испытали однажды, когда придумали хитрый способ взломать эту игру. Нужно было пробить сквозь кирпичики на краю экрана туннель; тогда, если шарик пролетал сквозь него и оказывался сверху от стены, он начинал скакать, отражаясь от верхней кромки экрана и верхних, «дорогих», кирпичиков и постепенно уничтожая стену. Игрок мог расслабиться и просто смотреть за этим процессом, пока шарик в конце концов не прилетал сквозь стену обратно вниз. Нужно было только держать ракетку наготове, чтобы снова отбить шарик вверх. Очень приятная была стратегия!

Хассабис и другие члены группы, которую он собирал в это время, также много играли в молодости в компьютерные игры. Возможно, их родителям было приятно узнать, что время и силы, потраченные на эти игры, не были растрacены впустую. Игра Breakout оказалась идеальным полигоном для проверки способности коллектива DeepMind запрограммировать компьютер на обучение играм. Написать программу для каждой отдельной игры было бы работой сравнительно несложной. Но Хассабис и его коллеги ставили перед собой гораздо более трудную задачу.

Они хотели написать программу, которая получала бы на входе состояние пикселей экрана и текущий счет и играла так, чтобы максимизировать счет. Правила игры программе не сообщаются: она должна случайным образом экспериментировать, двигая в разные стороны ракетку в Breakout или по-всякому стреляя из лазерной пушки по снижающимся кораблям пришельцев в Space Invaders. Каждый раз, когда программа делает ход, она может оценить, привел ли он к увеличению счета или не произвел никакого эффекта.

Эта программа реализует возникшую еще в 1990-е годы концепцию обучения с подкреплением, которая предполагает корректировку вероятности определенных действий в зависимости от воздействия на функцию вознаграждения или счет. Например, в игре Breakout можно принять только одно-единственное решение – сдвинуть ракетку, расположенную внизу экрана, влево или вправо. Изначально выбор делается с вероятностью 50: 50. Но если случайное перемещение ракетки приводит к попаданию по шарик, то через короткое время после этого счет увеличивается. Тогда программа производит перекалибровку вероятности смещения влево или вправо с учетом этой новой информации. Это увеличивает вероятность смещения в том же направлении, в котором движется шарик. Новшеством было совмещение такого обучения с нейронными сетями, которые должны были оценивать состояние пикселей и решать, какие именно элементы коррелируют с ростом счета.

Вначале, поскольку компьютер просто пробовал случайные ходы, его игра была ужасна; он почти не набирал очков. Но каждый раз, когда очередной случайный ход приводил к увеличению счета, программа запоминала этот ход и более активно использовала его в дальнейшем. Постепенно случайные ходы прекратились, и стал проявляться более обоснованный рисунок игры – ходы, которые, как программа выяснила на опыте, по-видимому, способствовали росту счета.

Видеоролик, который сотрудники DeepMind приложили в качестве иллюстрации к написанной впоследствии статье об этой работе, стоит посмотреть. В нем показано, как программа учится играть в Breakout. Сначала видно, как она случайным образом двигает ракетку взад и вперед, чтобы посмотреть, что из этого выйдет. Затем, когда шарик наконец попадает в ракетку, отскакивает от нее и разбивает кирпичик, отчего увеличивается счет, программа начинает переписывать самое себя. По-видимому, соприкосновение пикселей ракетки с пикселями шарика приносит положительный результат. После 400 партий программа играет уже по-настоящему сильно: ракетка постоянно отправляет шарик то туда, то сюда.

Но настоящий шок я испытал, когда увидел, что она открыла, сыграв 600 партий. Она нашла нашу лазейку! Не знаю точно, сколько партий потребовалось сыграть нам в детстве, чтобы освоить этот фокус, но, судя по количеству времени, которое потратили мы с другом, их вполне могло быть и больше. И вот тебе на. Манипулируя ракеткой, программа стала про-

бывать по бокам туннели, чтобы шарик застревал в пространстве между верхом стены и верхним краем экрана. После этого счет начинает расти очень быстро, а компьютеру почти ничего не приходится делать. Насколько я помню, когда мы с моим другом открыли этот трюк, мы запрыгали от восторга. Машина же не ощутила ничего.

К 2014 году, через четыре года после создания DeepMind, программа научилась обыгрывать человека в двадцати девяти из сорока девяти игр для Atari, которые были ей предложены. Статья с подробным описанием достижений группы была опубликована в журнале Nature в начале 2015 года. Любая публикация в Nature считается одной из вершин карьеры ученого. Но эта статья удостоилась еще более высокой чести: она стала главной темой всего выпуска. Редакция журнала признала ее появление важным этапом в развитии искусственного интеллекта.

Стоит еще раз подчеркнуть, насколько поразительным было это достижение с точки зрения программирования. Программа, имевшая в своем распоряжении только сырые данные о состоянии пикселей и изменениях счета, прошла весь путь от случайных перемещений ракетки Breakout взад и вперед к пониманию того, что создание туннеля на краю стены позволяет добиться максимального счета. Но игры для Atari – далеко не ровня древней игре го. Хассабис и его коллеги по DeepMind решили, что готовы создать новую программу, которая могла бы взяться и за эту задачу.

Именно тогда Хассабис решил продать свою компанию Google. «Мы не собирались этого делать, но в течение трех лет я был настолько сосредоточен на поисках финансирования, что на исследования оставалось всего 10 % моего времени, – объяснял он в то время в интервью журналу Wired. – Я понял, что в одной жизни, наверное, не хватит времени и построить компанию размером с Google, и решить задачу создания ИИ. Чем я буду больше гордиться потом – созданием многомиллиардного дела или вкладом в раскрытие тайны разума? Выбор был очевиден». Благодаря этой продаже он получил в свое распоряжение всю мощь Google и возможность работать над созданием кода для достижения своей цели – разрешения проблемы го... а там и интеллекта.

Первая кровь

Преыдушие компьютерные программы, созданные для игры в го, не могли и приблизиться к уровню, достаточному для игры на равных даже против сильного любителя. Поэтому многие эксперты относились к мечте коллектива DeepMind создать программу, способную хотя бы отдаленно сравниться с мастерами международного класса, в высшей степени скептически. Большинство по-прежнему было согласно с мнением, которое высказал в газете New York Times в 1997 году, после шахматной победы компьютера DeepBlue, астрофизик Пит Хат: «До того как компьютер обыграет человека в го, пройдет, возможно, еще лет сто – а может быть, и больше. Любой более или менее умный человек, научившийся играть в го, через несколько месяцев сможет победить все существующие компьютерные программы. Для этого не надо быть Каспаровым».

Из этого столетия прошло всего двадцать лет, когда сотрудники DeepMind решили, что им, возможно, удалось решить эту задачу. Казалось, что их стратегия, по которой алгоритм должен был учиться и приспосабливаться, работает, но они не знали точно, насколько сильным получался этот алгоритм. Поэтому в октябре 2015 года они решили испытать свою программу в негласном состязании с чемпионом Европы того времени, родившимся в Китае игроком по имени Фань Хуэй.

Программа AlphaGo разгромила Фань Хуэя, выиграв все пять партий из пяти. Но между европейскими и дальневосточными игроками в го существует огромный разрыв. Лучшие из европейских игроков занимают места всего лишь в шестой сотне мирового рейтинга. Поэтому, хотя такая победа была впечатляющим достижением, ее можно было сравнить с результатом

испытаний беспилотного автомобиля, который сумел обогнать на трассе Сильверстоун человека за рулем «форда-фиесты», а потом пытается состязаться с Люисом Хэмилтоном в гонке «Формулы-1».

Во всяком случае, когда пресса дальневосточных стран узнала о поражении Фань Хуэя, она безжалостно и презрительно рассуждала о том, насколько неважной была победа AlphaGo. Когда появились известия о матче, жена Фань Хуэя даже звонила ему в Лондон и просила не выходить в интернет. Нечего и говорить, что он не устоял перед искушением. Чтение высокомерных отзывов комментаторов из родной страны, рассуждавших, что он не годится в достойные противники AlphaGo, было, разумеется, не слишком приятно.

Фань Хуэй утверждает, что вынес из матчей с AlphaGo новые идеи относительно игры. В следующие месяцы его рейтинг вырос: если до этого он занимал 633-е место, то теперь оказался в четвертой сотне. Но учился не только Фань Хуэй. Каждая партия, сыгранная AlphaGo, воздействует на программный код и изменяет его так, чтобы в следующей партии программа играла еще лучше.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

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

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