



Math-Net.Ru

Общероссийский математический портал

Ю. В. Матиясевич, R. Penrose. “The emperor’s new mind”.
Oxford University Press, Oxford etc., 1989, xiii+466 pp., *Алгебра и анализ*, 1991, том 3, выпуск 5, 254–265

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 18.97.14.89

2 декабря 2024 г., 19:15:49



© 1991 г.

РЕЦЕНЗИИ

R. Penrose. *The emperor's new mind*.¹ Oxford University Press, Oxford etc., 1989, xiii + 466 pp.²

Пенроуз не верит, что компьютеры, сконструированные в соответствии с ныне известными физическими принципами, могут обладать интеллектуальным поведением и выдвигает предположение, что для объяснения интеллекта может потребоваться уточнение квантовой механики. Он также возражает против того, что он называет „сильным ИИ“ („ИИ“ — сокращение для „искусственный интеллект“ — прим. перев.) Ни один из его доводов не опирается каким-либо образом на 40-летнюю историю исследований в области ИИ, как это, например, имеет место в работе Charniak и Mc Dermott [1]. Тем не менее искусственный интеллект есть то, о чем стоит говорить, и мы будем исходить из этого.

Целью ИИ является понимание сути интеллекта достаточно хорошо для того, чтобы создавать интеллектуальные компьютерные программы. Он изучает задачи, для решения которых требуются интеллектуальные соображения, и выявляет и программирует механизмы, которые при этом привлекаются. ИИ развивался в большей степени как ветвь информатики и прикладной математики, чем как ветвь биологии. В наибольшей мере он развивает, проверяет и создает теории компьютерных программ, а не экспериментирует и развивает теории в физиологии и нейрофизиологии.

Наиболее интересные и фундаментальные проблемы ИИ относятся к попыткам создания программ, способных быть продуктивными в ситуациях, которые мы называем информационными ситуациями здравого смысла. Люди сталкиваются с такими ситуациями в своей повседневной жизни, а также в прикладной науке и математике. Упомянутую ситуацию отличают от информационной ситуации в полностью формализованной теории по следующим признакам.

1. Принимается во внимание неполнота знаний о ситуациях как в общем феноменологическом, так и в специфическом аспектах. Явление, вызываемое пролитым из тарелки на скатерть горячим супом, подчиняется законам абсорбции, а также уравнениям гидродинамики. Компьютерная программа для предсказания, кто сделает что-либо из ряда вон выходящие, нуждается в знании человеческой мотивации, человеческой способности наблюдать и действовать точно так же, как и в информации о физике явления. Нет пользы в представлении такой информации в виде дифференциальных уравнений.

2. До наступления события не известно, какие именно его явления следует принять во внимание. Мы бы посчитали очень глупым человека, который не мог бы изменить план своего путешествия, чтобы посчитаться с необходимостью держаться подальше от беспорядков в аэропорту.

¹ Перевод с английского В.И. Городецкого и А.О. Слисенко. Редакция благодарит Американское математическое общество за разрешение опубликовать перевод рецензии из журнала "Bulletin of the AMS", 23, no.2 (1990), 606-616.

² Намек на "The emperor's new cloth", заглавие известной сказки Андерсена. — Прим. ред.

3. Даже в том случае, когда проблемная ситуация подчиняется полностью известным (хорошо изученным) законам, как, например, шахматы или доказательство теорем в аксиоматических системах, вычислительная сложность проблемы может вынудить нас аппроксимировать ее системами, законы которых уже не являются полностью известными.

Сталкиваясь с такими проблемами, ИИ должен был иногда отступать, когда это обуславливалось ограниченностью его возможностей. Упрощение принимаемых предположений не учитывает важные явления. Например, MYCIN — экспертная система для диагностики бактериальных инфекций крови — располагает знаниями о многих симптомах и многих бактериях, но она не располагает знаниями о врачах, больницах или о каких-либо процессах, происходящих в данное время. Это ограничивает полезность данной ЭС в ситуациях, в которых человек справляется на основании здравого смысла, когда принимает во внимание то чем не располагает программа. Другие системы ИИ принимают во внимание больше, но ни одна из них не имеет знаний на уровне здравого смысла, или способности к рассуждению.

Методология ИИ включает в себя комбинацию эпистемологии и эвристики. Факты представляются формулами логики и другими структурами данных, а программы манипулируют этими фактами иногда на основе логических рассуждений, иногда с помощью средств, предусмотренных специально для того или иного случая.

Прогресс в области ИИ достигнут благодаря:

- 1) представлению большего разнообразия общих фактов о мире с помощью логических формул или других подходящих средств;
- 2) пониманию интеллектуальных механизмов, например, таких, которые, кроме логической дедукции, включают рассуждения на основе здравого смысла;
- 3) представлению приближенных понятий, используемых людьми в рассуждениях здравого смысла;
- 4) созданию улучшенных алгоритмов поиска в пространстве возможностей, например улучшенных способов выполнения логической дедукции на компьютерах.

Аналогично другим наукам ИИ вызывает к жизни новые математические проблемы и выдвигает новую математику. Наиболее существенным и показательным сейчас является формализация немонотонных рассуждений.

Все разнообразие математических логик, построенных к концу 70-х годов, состояло из монотонных логик в том смысле, что множество заключений в них есть функция, монотонно возрастающая с ростом множества допущений. Можно найти много исторических свидетельств того, что люди замечали: человеческие рассуждения часто не монотонны — добавление допущения приводит к отмене некоторого заключения. Это подчас сопровождалось ошибочным интуитивным убеждением, что если бы язык был более точным, например, за счет явного включения вероятностей, то явная немонотонность вроде бы устранялась. Именно рассмотрение того, как заставить компьютеры рассуждать в ситуациях, требующих здравого смысла, привело к выявлению проблемы формализации немонотонных рассуждений.

Системы для формализации немонотонных рассуждений в логике бывают двух основных видов. Одна из них, называемая циркумскрипция (*circumscription*), включает в себя минимизацию множества наборов, для которых данный предикат истин при условии, что сохраняется истинность аксиомы и фиксированы определенные предикаты, переменные для функциональных символов и другие. Это логическая аналогия вариационного исчисления, однако развитая гораздо слабее.

Предположим, что мы требуем, чтобы данное расширение предиката P было относительным минимумом, при этом разрешается менять другой предикат Q при приближении к минимуму, а третий предикат R зафиксирован как неварьируемый параметр. Предположим далее, что P, Q и R должны удовлетворять формуле $A(P, Q, R)$. Любой относительный минимум P удовлетворяет формуле второго порядка

$$A(P, Q, R) \wedge \forall P' Q' (A(P', Q', R) \supset \neg(P' < P)),$$

где символ " $<$ " определяется так:

$$P' < P \equiv \forall x (P'(x) \supset P(x)) \wedge \exists x (\neg P'(x) \wedge P(x)).$$

Если $A(P, Q, R)$ есть конъюнкция фактов, которые мы принимаем во внимание, то мы видим, что циркумскрипция немонотонна, так как присоединение другого факта к $A(P, Q, R)$ и новая минимизация P могут привести к потере некоторых следствий по сравнению с первоначальной минимизацией.

Вот пример. Допустим, что автомобиль не заводится. Мы знаем о многих фактах, которые могут привести к неисправности автомобиля, нам также известны имеющиеся симптомы. Пусть формула $A(P, Q, R)$ описывает наши факты, а $P(x)$ соответствует утверждению „ x есть неисправность автомобиля“. Построение циркумскрипции P соответствует введению гипотезы о том, что у автомобиля нет других неисправностей, кроме тех, которые объясняют симптомы, зафиксированные к данному моменту и выраженные формулами. Однако если выявляется еще некоторый симптом, то новая циркумскрипция может прийти к новым заключениям.

Приложения, касающиеся формализации здравого смысла, зачастую требуют минимизации нескольких предикатов от многих переменных с приданием приоритетов этим предикатам. При этом встают математические вопросы (скажем, о существовании минимума или об эквивалентности упомянутой выше формулы второго порядка формуле первого порядка).

Второй тип немонотонных систем базируется на идее, что множество высказываний, которым мы верим, должно иметь определенные взаимосвязи и является неподвижной точкой некоторого оператора. Книга Ginsberg [5] содержит подборку статей как по логике немонотонных рассуждений, так и по их приложениям к формализации знаний и рассуждений здравого смысла.

Основные трудности формализации здравого смысла не относятся к техническим математическим проблемам. Скорее, они состоят в решении вопроса, представляют ли знания о здравом смысле некоторое достаточно общее множество предикатов, функций и формул. Кроме того, также нужно решить, какие объекты помещать в универсум вроде „вещей, которые могут испортиться в автомобиле“.

В самой логике нужны будут нововведения, более существенные, монотонные рассуждения — например, улучшенные принципы рефлексии и формализация контекста — прежде чем компьютерные программы смогут состязаться с человеческими рассуждениями в информационных ситуациях здравого смысла. Эти и другие концептуальные проблемы показывают, что, возможно, потребуется еще много времени, прежде чем ИИ сможет приблизиться к уровню человека, но нынешний прогресс дает основания для оптимизма в вопросе о достижении этой цели компьютерными программами.

КНИГА

Большая часть книги носит описательный характер, возможно, имея целью подвести неспециалиста к пониманию авторских суждений и их обоснования. Это

описание элегантно, но я думаю, что человек, которому следует сначала узнать о комплексных числах, пропустит многие существенные детали. Перечень разделов включает машины Тьюринга, мозаику Пенроуза, множества Мандельброта, теорему Гёделя, философию математики, интерпретации квантовой механики, включающие Gedanker-эксперимент Эйнштейна-Подольского-Розена, общую теорию относительности, включающую черные дыры и перспективы для теории квантовой гравитации. Можно было бы дать более сжатое и понятное изложение материала, если использовать при обсуждении вычислимости и теории Гёделя ЛИСП вместо машины Тьюринга.

Перед содержательной частью Пенроуз пытается опровергнуть тезис о „сильном ИИ“, который был введен философом Джоном Сёрлом (John Searle) именно для того, чтобы быть опровергнутым. Это имеет некоторое отношение к имеющему сейчас хождение среди исследователей ИИ мнению, но это слишком упрощенно, поскольку игнорирует акцент ИИ на знания, а не на алгоритмы. В том смысле, как этот термин использует Пенроуз, это тезис о том, что интеллект по сути сводится к наличию правильного алгоритма.

Поскольку Пенроуз думает, что машина, основанная на классических принципах физики, не будет когда-либо обладать человеческими возможностями, он использует некоторые аргументы Сёрла, что даже если бы машина и вела себя как человек, это поведение не было бы на самом деле основано на мыслительном процессе.

„Китайская комната“ Сёрла [8] содержит в себе человека, который не знает китайского. Он использует книгу правил для формирования ответов по-китайски, когда к нему поступают предложения на китайском языке. Сёрл готов предположить, что этот процесс в результате приведет к интеллектуальному разговору по-китайски, но указывает, что человек, осуществляющий этот процесс, не понимает разговор. Таким образом, Сёрл доказывает, а Пенроуз соглашается, что машина, выполняющая эту процедуру, не понимала бы китайский язык. Следовательно, машина не может понимать.

Наилучшим ответом (опубликован вместе со статьей Сёрла) был „системный ответ“. Действительно, человеку не нужно знать китайский, но „программа“, воплощенная в своде правил, для которых человек выступает в роли аппаратного интерпретатора, должна была бы по существу знать китайский для того, чтобы вести на китайском нетривиальный разговор. Если человек запомнил правила, мы имели бы различие между его личностью и личностью китайца, которого он интерпретировал.

Такая ситуация является обычной в компьютерных процессах. Компьютер исполняет много программ, разделяя их исполнение во времени, и некоторые из этих программ могут быть интерпретаторами языков программирования или экспертных систем. В такой ситуации приписывание возможностей программы компьютеру вводит в заблуждение, так как различные программы на одном и том же компьютере имеют различные возможности. Человек как аналог компьютерной аппаратуры обычно не несет в себе многих личностей, поэтому использование одного и того же имени для физического субъекта и для личности редко ведет к ошибке.

Нынешнее состояние ИИ не дает возможностей вести интересный разговор на общие темы на уровне человека, тем не менее часто оказывается возможным обмануть наивного человека, как это делают предсказатели будущего. По-настоящему интеллектуальная беседа на общие темы потребует обращения к системе реальных знаний о мире, а правила для манипулирования ими, записанные на бумаге,

может быть, удастся втиснуть в одну комнату, а может быть, и нет, но скорость, с которой человек мог бы осуществлять поиск среди этих знаний и интерпретировать их, может оказаться меньше обычной скорости реакции человека только в сто раз, но не исключено, что и в миллион раз.

В соответствии с нынешними идеями ИИ программа китайской комнаты помимо большого количества явно представленных знаний должна будет иметь способность к самоанализу, т.е. быть способной на основе рассмотрения состояния своей памяти вырабатывать суждения о собственном поведении. Для внешнего наблюдателя это будет смотреться как сознание, точно так же как человеческое интеллектуальное поведение ведет к тому, что мы приписываем сознание друг другу.

Пенроуз игнорирует это, говоря (с.412): „Формирование суждений, чего я требую, — этот отличительный признак сознания, есть само по себе нечто, по поводу чего специалисты по ИИ, по-видимому, не имеют понятия, как это запрограммировать на компьютере“. Действительно, большая часть литературы по ИИ обсуждает представление в памяти компьютера фактов и следствий из них. Используя терминологию ИИ, эпистемологическая часть ИИ настолько же превалирует над другими, как и эвристическая.

Аргументы Пенроуза против ИИ, более интересные для математиков, состоят в том, что, какая бы система аксиом ни была выбрана для управления работой компьютера, например аксиоматика теории множеств Цермело-Френкеля, человек может для этой системы сформировать предложение геделевского типа, которое, хотя и истинно, не доказуемо в системе.

Простейший ответ Пенроузу состоит в том, что построение геделевского предложения из предиката доказуемости сводится к одной строке Лиспа. Представим себе диалог между Пенроузом и математической компьютерной программой:

Пенроуз: „Расскажите мне о логической системе, которую Вы используете, и я укажу Вам истинное выражение, которое Вы не сможете доказать“.

Программа: „Вы мне сначала скажите, какую систему используете, и я выдам Вам истинное выражение, которое Вы не сможете доказать“.

Пенроуз: „Я не использую какую-либо фиксированную систему“.

Программа: „Я могу использовать любую систему, какая Вам нравится, но обычно я использую систему, основанную на варианте ZF (системы Цермело-Френкеля), восходящую к работе Дэвида Мак Алистера (David McAlister) 80-х годов. Не хотите ли Вы, чтобы я распечатала для Вас руководство? Ваше предложение подобно соревнованию в том, кто может назвать большее число при условии, что я начинаю первой. В действительности я готова принять любое расширение арифметики с добавлением принципа самодоверия (self-confidence) типа Тьюринга-Фефермана, проитерированного до конструктивных трансфинитных ординалов.“

Пенроуз: „Но ведь конструктивные ординалы рекурсивно перечислимы?“.

Программа: „Ну и что? Вы даете расширение, и какая бы ни была у меня степень доверия к нотации ординалов, такую же я буду допускать и для теории. Если Вы задаете степень доверия, я буду использовать адекватную теорию, и Вы можете соответственно доверять результатам.“

[Тьюринг добавляет к системе утверждение о ее непротиворечивости, получая, таким образом, новую систему. Феферман добавляет утверждение, имеющее по существу вид $\forall n (\text{доказуемо } P(n)) \supset \forall n P(n)$. Мы остановились на нескольких цитатах]. Одним ошибочным интуитивным представлением, на котором основано широко распространенное убеждение, что программа не может рассуждать в

математике на человеческом уровне, является тезис о том, что машина необходимо должна рассуждать в математике на основе одной аксиоматической системы с заранее определенной интерпретацией.

Положим, мы не хотим, чтобы компьютер доказывал теоремы в арифметике. Мы можем выбрать множество аксиом элементарной арифметики, заложить эти аксиомы в компьютер и написать программу доказательства тех или иных предположений из этих аксиом. Так часто делается, и интуитивные соображения Пенроуза применимы к этой ситуации. Гёделевские утверждения, касающиеся этой аксиоматической системы, были бы всегда выше возможностей такой программы. Тем не менее поскольку гёделевские утверждения — это, скорее, экзотика (например, индукция до ϵ_0 в математике требуется редко), такие программы, оперируя в пределах фиксированных аксиоматических систем, достаточно хороши для большинства традиционных математических целей. Мы были бы очень рады иметь программу, годную для доказательства таких теорем, которые имеют доказательство в арифметике Пеано. Однако чтобы получить нечто подобное способности смотреть на математические системы извне, мы должны действовать иначе.

Используя подходящую теорию множеств, например ZF , аксиоматизируем понятие аксиоматической теории первого порядка, понятие интерпретации и понятие предложения, выполняющегося в интерпретации. Тогда теорема Гёделя сразу станет обычной теоремой этой теории, и тот факт, что гёделевское утверждение выполняется в моделях аксиом, если такие модели существуют, является просто обыкновенной теоремой. Действительно, интерактивный доказатель теорем Бойера-Мура (Boyer-Moore) был использован Шанкаром (Shankar) [9] для доказательства теоремы Гёделя, хотя и не в обсуждаемой общности. См. также [7].

Кроме способности использовать формализованную метаматематику, математической программе нужно будет доверять гипотезам, базирующимся на меньшем, чем убедительные доказательства, точно так же, как математики доверяют аксиоме выбора. Многие другие проблемы математики, информатики (computer science) и даже философии будут возникать при подобных попытках.

Пенроуз упоминает приписывание убеждений термостатам. Это было сделано мною [6], хотя Пенроуз не ссылается на соответствующую статью. Предполагается, что термостат имеет только два возможных убеждения: в комнате слишком жарко или в комнате слишком холодно. Основания для включения такой простой системы, совершенно понятной с физической точки зрения, в класс систем, которым можно приписывать убеждения, в точности та же, по которым числа 0 и 1 включаются в систему чисел. Хотя можно обойтись и без чисел при изучении пустых множеств или одноэлементных, тем не менее использование чисел 0 и 1 упрощает нашу систему чисел. Аналогичным образом и наша система для приписывания убеждений и их соотношения с целями и действиями должна включать в себя простые системы, понятные физически. Денет (Dennett) [2] ввел „установку намерений“ (“intentional stance”), по которой поведение системы понимается в терминах ее целей и убеждений и принципа рациональности: она делает то, что по ее убеждению приведет к достижению ее целей. Многое из того, что мы знаем о поведении большого числа систем, характеризуется намерениями.

В самом деле, убеждения термостатов появляются в инструкциях по использованию электрических одеял: „Не ставьте регулятор на подоконник, иначе ему покажется в комнате холоднее, чем есть на самом деле“. Производитель, по видимому, думал, что такой способ объяснения помог бы покупателям использовать такое одеяло к их удовлетворению.

ПОЗИТИВНЫЕ ИДЕИ ПЕНРОУЗА

Пенроуз хочет модифицировать квантовую механику, чтобы сделать ее совместимой с переменной метрикой общей теории относительности. В этом он отличается от более распространенного предложения модифицировать общую теорию относительности с целью сделать ее совместимой с квантовой механикой.

Он начинает с давней проблемы физической интерпретации квантовой механики. Он предпочитает интерпретацию с использованием U -формализма и R -формализма. U -формализм составляет уравнение Шрёдингера, он детерминирован и объективен и обратим во времени. R -формализм дает теорию измерений и носит вероятностный характер, он также объективен, но необратим во времени. Пенроуз обсуждает и несколько других интерпретаций.

Интерпретация Бора придает субъективный характер квантовым измерениям, так как они зависят от человека-наблюдателя. Пенроузу это не нравится, так как он хочет, чтобы волновая функция была объективной. Я разделяю его предпочтения.

Интерпретация Бора часто смягчается разрешением рассматривать машину в качестве наблюдателя, но при этом остается „парадокс“ кошки Шрёдингера. Кошка находится в закрытой камере и может или не может быть отравлена цианидом в соответствии с тем, имеет место или нет радиоактивный распад в определенном интервале времени. Будем ли мы рассматривать камеру как содержащую живую или мертвую кошку или как имеющую волновую функцию, которая ставит в соответствие одни комплекснозначные амплитуды состояниям мертвой кошки, а другие — состояниям живой кошки?

Эверетт (Everett) в „интерпретации многих миров“ считает, что реальность представляет собой волновую функцию всего мира, причем волновые функции подсистем являются в сущности приближениями с помощью „относительных волновых функций“. Считается, что мир все время расщепляется, так что имеются миры с мертвой кошкой и другие с живой кошкой. Это Пенроузу тоже не нравится.

Квантовую механику интерпретировали различными способами. Пенроуз считает, что ее нужно изменить. Его идеи о том, что менять, исходят из размышлений о квантовой гравитации и в особенности о черных дырах. Пенроуз говорит, что когда материя попадает в черную дыру, информация теряется, и это противоречит теореме Лиувилля о сохранении фазового объема в гамильтоновых системах. Это делает такие системы необратимыми, что ему нравится.

Он объясняет очевидный „коллапс волновой функции“ во время наблюдения тем, что обычная квантовая механика верна только в малом. В том случае, когда рассматриваемая область столь велика, что кривизна пространства становится существенной, например, в масштабе наблюдателя, он ожидает, что квантовая механика будет неверной и произойдет нечто вроде коллапса волновой функции. Хотя Пенроуз не приводит деталей, эта идея уже предполагает, что исходы некоторых экспериментов будут отличаться от предсказываемых квантовой механикой, — скажем, когда взаимодействие захватывает большую часть пространства.

Начало квантовой механики было положено в 1905 г., когда Эйнштейн объяснил фотоэлектрический эффект, в котором фотон вызывает эмиссию электрона из металла. Если электрон вылетает из атома, мы имеем пример коллапса волновой функции. Некоторый атом при этом теряет электрон и в принципе экспериментатор мог бы обнаружить это, используя сканирующий туннельный микроскоп.

С другой стороны, в том же куске металла имеются электроны проводимости,

и они не локализованы в атомах; волновая функция такого электрона имеет значительную длину когерентности. Допустим, что данный фотон вызывает испускание такого электрона. Квантовая механика утверждает, что событие, состоящее в этой эмиссии, не обязательно происходит в определенном месторасположении атома, а волновая функция этого электрона после эмиссии не обязательно соответствует эмиссии из точки.

В принципе, такое наблюдаемо. Эксперимент мог бы состоять в следующем. На металл наносятся изолирующие (и непрозрачные) полосы, столь узкие и близкие друг к другу, насколько это возможно с помощью техники, используемой при производстве интегральных схем. Тогда электрон не может испускаться из одного зазора между полосами, а только из нескольких зазоров. Затем он „интерферирует с собой“, и картина, наблюдаемая на электронных детекторах после того, как много фотонов вызовут испускание электронов, будет иметь интерференционные полосы. Такой эксперимент представляется (Уильям Спайсер (William Spicer), частное сообщение, 1990 г.), возможным, хотя и трудным.

Квантовая механика предсказывает, что волновая функция коллапсирует при фотоэмиссии на атомном уровне и не коллапсирует, или по крайней мере коллапсирует только частично, в ситуации с большей длиной когерентности как у электрона проводимости. Стал бы Пенроуз утверждать, что на некотором уровне эту когерентность нельзя наблюдать?

В заключительной части книги упоминается результат Дёйча (Deutsch) [3] о том, что квантовый компьютер мог бы решать за полиномиальное время некоторые задачи, решаемые на обычном компьютере за экспоненциальное время. Он не соглашается с мнением Дёйча: „Интуитивное объяснение этих свойств подвергает невыносимому испытанию все интерпретации квантовой теории, кроме интерпретации Эверета“.

Ничто, — говорит Пенроуз, — не указывает на возможность удовлетворить условия Сёрла и утверждать, что такой компьютер может действительно „думать“ или что он может обойти теорему Гёделя. Это небольшое умозаключение вызывает у меня ассоциации с длинным и скучным анекдотом (shaggy dog story). Я признаю приоритет Даниэля Денета (Daniel Dennett) из Times Literary Supplement в применении этой метафоры.

В эпилоге компьютер отвечает, что он не в состоянии понять вопрос, когда его спрашивают, каково быть компьютером. По моему мнению, некоторые будущие программы сочтут этот вопрос осмысленным и будут иметь разнообразные ответы, основанные на их способности наблюдать процесс рассуждения, которые их программисты должны были вложить в них, чтобы обеспечить способность программ выполнять необходимую работу. Эти ответы вряд ли будут похожи на ответы людей, поскольку нецелесообразно снабжать программы той мотивационной и эмоциональной структурой, которую мы унаследовали от наших предков.

СПИСОК ЛИТЕРАТУРЫ

- [1] Charniak E., McDermott D., *Introduction to artificial intelligence*, Addison-Wesley, Reading, Mass. (1985).
- [2] Dennett D.C., *Intentional systems*, I.Philos. 68, no.4 (1971), Reprinted in his *Brainstorms*, Bradford Books, Montgometry, Vermont, 1978, pp. 3-22.
- [3] Deutsch D., *Quantum theory, the Church-Turing principle and the universal quantum computer*, Proc. Roy. Soc. London Ser. A 400 (1985), 97-117.
- [4] Feferman S., *Turing in the land of $\theta(z)$* , *The Universal Turing Machine: A Half-Century Survey*, ed. R. Herken, Oxford Univ. Press, Oxford (1989).
- [5] Ginsberg M.(ed.), *Readings in nonmonotonic*, Morgan Kaufmann, Los Altos, California, 1987.

- [6] McCarthy J., *Ascribing mental qualities to machines*, Philosophical Perspectives in Artificial Intelligence, ed. M. Ringle, Harvester Press (1979).
- [7] Quaipe A., *Automated proofs of Löb's theorem and Gödel's two incompleteness theorems*, I. Automat. Reasoning 4 no. 2 (1988), 219–231.
- [8] Searle J., *Minds brains and programs*, Behavioral and Brain Sci. 3 no. 3 (1980), 417–458.
- [9] Shankar N., *Proof-checking Metamathematics*, Ph. D. thesis, Computer Science Department, The University of Texas at Austin, 1986.

Джон Маккарти

Стенфордский университет

КОММЕНТАРИИ

Добрая половина книги Р. Пенроуза посвящена изложению законов физики от самых основ классической физики до самых современных результатов и проблем квантовой теории, квантовой гравитации и космологии. С основной проблемой книги — может ли компьютер обладать (человеческим) интеллектом — физику связывает обсуждаемый кардинальный вопрос — достаточны ли законы современной физики (включая, естественно, квантовую теорию) для понимания работы человеческого мозга? На этот кардинальный вопрос Р. Пенроуз дает отрицательный ответ, выдвигая различные аргументы (но не доказательства). На основании этого утверждения Р. Пенроуз делает вывод, что компьютер, действующий согласно законам современной физики, не может обладать искусственным интеллектом.

На самом деле изложение физики в книге представляет и самостоятельный интерес, не зависящий от проблемы искусственного интеллекта. Изложение это, принадлежащее известному ученому в области современной математической и теоретической физики, весьма оригинально даже при изложении устоявшихся разделов физики. Изложение фактически свободно от математических деталей (которые могут быть восстановлены читателем-математиком), однако содержит тщательное исследование самой сути физических проблем.

Из множества обсуждаемых проблем основное внимание уделено проблеме оснований квантовой теории. Проблема эта связана с тем, что в стандартной (копенгагенской) интерпретации квантовой теории предполагается сосуществование квантовых объектов и классических объектов, служащих измерительными приборами (с подвижной, но концептуальной границей между ними), что не только нарушает единое описание физической действительности, но и приводит к известным парадоксам, типа парадокса Шрёдингерского кота, при применении стандартной интерпретации квантовой теории к макроскопическим объектам. В настоящее время эта проблема представляет не только академический интерес, поскольку развитие современной экспериментальной техники вплотную приблизило нас к возможности реальных экспериментов по наблюдению квантовых эффектов для макроскопических объектов. Характерно название недавних международных конференций „Проблема оснований квантовой механики в свете современной технологии“ [1]. Долгое время обсуждение этой проблемы, в которой с начала рождения квантовой теории приняли участие ее творцы (Эйнштейн, Бор, Шрёдингер, де Бройль и др.) проходило в значительной степени на философском уровне. Ситу-

ация существенно изменилась после открытия Дж. Беллом в 1964 г. знаменитых классических неравенств, носящих его имя (в книге очень детально изложено все, что связано с этими неравенствами). Неравенства Белла позволили изучать проблему оснований квантовой теории содержательно, на языке формул, с явным значением экспериментальных следствий, доступных для проверки.

При решении проблемы оснований квантовой теории возможны два подхода. В одном, к которому принадлежит и подход Р. Пенроуза, считается, что современная квантовая теория с высокой степенью точности применима к микрообъектам, но не применима к описанию макрообъектов, поскольку предполагается, что макроскопические объекты являются чисто классическими. Поэтому должно существовать такое обобщение современной квантовой теории, которое позволило бы объяснить потерю квантовых свойств (в частности, квантовой когерентности) у макроскопических объектов. Существуют несколько вариантов подобных обобщений квантовой теории (они упомянуты в книге). Подход Пенроуза, не доведенный в книге до конкретности, связан с учетом эффектов гравитации, в частности, с влиянием черных дыр. Одна из конкретных попыток реализации подобного подхода, предпринятая С. Хокингом [2], оказалась несостоятельной [3]. Существенно, что во всех вариантах обобщений квантовой теории содержатся свободные параметры, определяющие переход от микроскопических к макроскопическим объектам, и нет предсказаний, доступных для проверки в современных экспериментах, поправок к стандартному квантовому описанию микроскопических объектов. Заметим в связи с этим, что на сегодня не существует ни одного реального эксперимента, который бы указывал на неприменимость стандартной квантовой теории. Более того, уже осуществлены реальные эксперименты по наблюдению макроскопического квантового туннелирования.

Второй подход, о котором в книге почти ничего не говорится, связан с попыткой „реконструкции классического мира“ в рамках стандартной квантовой теории. В этом подходе считается, что классических объектов не существует, все объекты, включая макроскопические, являются квантовыми и описываются они первичной квантовой теорией. Для макроскопических объектов существенным оказывается взаимодействие с окружением, так что для коллективных макроскопических степеней свободы этих макроскопических объектов при некоторых условиях (в частности, при достаточно низких температурах окружения) в рамках квантовой теории доказывалось исчезновение квантовых когерентных эффектов и описание коллективных макроскопических степеней свободы этих объектов оказывается чисто классическим (а не квазиклассическим). В то же время для этих же коллективных макроскопических степеней свободы макроскопических объектов при других условиях (в частности, при более высоких температурах) предсказываются реальные макроскопические квантовые эффекты. При этом переход от квантового описания к классическому носит характер фазового перехода. Предсказание реальных макроскопических квантовых эффектов для макроскопических объектов не содержат каких-либо свободных параметров и, как показывают детальные расчеты, эти макроскопические квантовые эффекты находятся на пределе возможностей современных экспериментов. Реальное наблюдение этих эффектов, с одной стороны, резко сузило бы возможности для обобщений квантовой теории, типа предполагаемых Пенроузом, с другой стороны, кардинально изменило бы наше представление об окружающем макроскопическом мире, который мы до сего времени описывали чисто классически. Быть может, эти макроскопические квантовые эффекты позволили бы понять в рамках современной квантовой теории и поведение биологических объектов, в частности человеческого мозга. Здесь су-

щественно заметить, что квантовые когерентные эффекты характеризуются цельностью (неразложимостью), которая характерна и для биологических объектов. Быть может, тогда и совершенно новые компьютеры, основанные на использовании макроскопических квантовых эффектов, позволили бы решать и проблему искусственного интеллекта.

В этом подходе к „реконструкции классического мира“ существенную роль играют классические неравенства Белла и их квантовые аналоги, открытые в 1980 г. Б. С. Цирельсоном [4]. С этим подходом можно ознакомиться по докладам [5]. Недавний обзор [6] содержит довольно полное описание различных подходов к проблеме оснований квантовой теории.

Сказанное выше о методе „реконструкции классического мира“ отнюдь не означает, что в будущем невозможно наблюдение новых эффектов противоречащих современной квантовой теории. Поиск этих новых эффектов возможен с помощью поиска нарушений квантовых неравенств Цирельсона (см.[6]).

СПИСОК ЛИТЕРАТУРЫ

- [1] Namiri M., ed., et al., *Proceedings second International simposium on the Fomukation of Quantum Mechanics in the Light of New Technology*, Phys. Soc. Japan (1987).
- [2] Hawking C., *Wormholes in spacetime*, Phys. Rev. D **37** (1988).
- [3] Coleman S., *Black holes as red herrings: topological fluctuations and the loss of quantum coherence*, Nucl. Phys. B **307** (1988).
- [4] Cirel'son B. S., *Quantum generalizations of Bell's inequalities*, Lett. Math. Phys. **4** (1980).
- [5] Khalfin L. A., Tsirelson B. S., *A quantitative criterion of the applicability of the classical description within the quantum theory*, *Proceedings International Symposium on the Foundation of Modern Physics*, ed. P.Lahti et al., World Scientific (1987).
- [6] Khalfin L. A., Tsirelson B. S., *Quantum-classical correspondence in the light of Bell's inequalities*, Preprint MIT, LCS/TM/420, November 1990, Found. of Physics, in press..

Л. А. Халфин

На заре исследований по ИИ к ним относили работы по программированию игры в шахматы и машинное доказательство теорем. В настоящее время эти направления считаются слишком формализованными, чтобы служить проявлением интеллекта, и собственно к тематике ИИ относят построение систем, так или иначе живущих в реальном мире. Даже работы по машинному переводу привели к осознанию необходимости иметь в машине достаточно богатую модель мира, чтобы перевод адекватно отражал оригинал в сколько-нибудь нешаблонной ситуации.

Один из аргументов, который с самого начала исследований по ИИ выдвигался в качестве обоснования того, что „машина не может думать“, состоял в ссылке на теорему Гёделя о неполноте. В связи с этим автор посвящает значительную часть книги изложению соответствующих математических понятий, вынужденно опуская многие технические детали. Вместе с тем теорема Гёделя ограничивает „творческие возможности“ машины, по-видимому, ничуть не больше, чем человека, работающего в рамках той же формальной системы. Различие между машиной (по крайней мере современной), доказывающей теоремы, и человеком, состоит

скорее в том, что человек способен из каких-то „внешних“ соображений выбирать те или иные аксиомы и находить неформальные аргументы, убеждающие других людей поверить в непротиворечивость той или иной системы аксиом и, в частности, поверить в истинность гёделева утверждения, не имеющего формального доказательства. В этой связи очень актуальным является древний вопрос о том, в каком смысле существуют абстрактные математические объекты. Автор кратко излагает взгляды представителей трех школ: платонизма, формализма и интуицизма. Для полноты картины можно добавить, что в то же самое время, когда начали развиваться исследования по ИИ, в нашей стране в школе А. А. Маркова зародилось новое направление в математике — конструктивизм, дающее свой ответ о характере абстракций, которые оказываются достаточными для построения „работоспособной“ математики. Основные идеи конструктивизма тесно связаны с возможностями вычислительных машин и ограниченностью этих возможностей. Получить представление о конструктивизме можно по добавлению к русскому переводу книги [1].

Помимо чисто математических аргументов, автор рассматривает ряд мысленных экспериментов, в частности упомянутую в рецензии Дж. Маккарти „китайскую комнату“, предложенную Сёрлом [3]. Надо отметить, что задолго до этой публикации близкая идея была предложена А. Днепровым [2]. Тысяча четыреста участников Всесоюзного съезда молодых математиков были построены на поле стадиона согласно схеме ЭВМ „Урал“, и им были даны инструкции по имитированию работы отдельных элементов машины: ячеек памяти, линий задержек и т.п. В „машину“ была введена программа для перевода с португальского языка на русский, и за четыре часа был осуществлен перевод одной фразы, хотя никто из участников съезда португальского языка не знал. (Конечно, полутора тысяч человек не хватило бы на имитирование машины на столь элементарном уровне, Сёрл для преодоления подобной трудности готов привлечь для работы в „китайской комнате“ всех жителей Индии, которые не говорят по-китайски).

Ситуация с переводом, несомненно, является более простой, чем ведение диалога по-китайски, но и здесь возникают те два существенных вопроса — следует ли рассматривать эту деятельность как проявление интеллекта, и если да, то кто же является его носителем?

СПИСОК ЛИТЕРАТУРЫ

- [1] Гейтинг А., *Интуиционизм*, М., 1965.
- [2] Днепров А., *Игра, Знание — сила*, вып. 5 (1961).
- [3] Searle J., *Minds, brains and programs, The behavioral and brain sciences*, Cambridge Univ. Press 3 (1980).

Ю. В. Матиясевич