

учной недостаточности» нашей теории (если здесь употребимо слово «теория») морали. В отличие от этих, претендующих считаться научными, работ, в публицистике можно встретить обращение к одной из рассматриваемых категорий — к категории высшего блага. Но в какой интерпретации? Утверждается, что если раньше высшей целью у нас признавалось благо общества, то теперь, наоборот, таковой надо считать благо личности. Говорить так, значит заменять одну крайнюю и одностороннюю точку зрения другой, столь же крайней и односторонней. Если первая означает превращение личности в «винтик» общественного механизма, ее угнетение в интересах командно-административной системы, «государственного класса», то вторая неминуемо ведет к теоретическому оправданию эгоизма и анархии, а практически — к уничтожению всякой общественности в конечном счете и к гибели самих индивидов, ибо жизнь людей невозможна вне общества. Ясно, что с помощью одной лишь категории высшей цели (высшего блага) основная проблема теории и практики морали — о соотношении личного и общественного — не может быть разрешена.

Огромная заслуга И. Канта в развитии этической мысли в том, что он поставил рядом с ней и глубоко разработал еще одну фундаментальную категорию — категорию конечной цели. Кант в определенной мере выявил диалектику высшей и конечной цели. Задача современной, в том числе советской, этики — усвоить и эффективно использовать это выдающееся достижение гениальной мысли великого философа.

¹ Аристотель. Никомахова этика // Соч.: В 4-х т. М.: Мысль, 1983. Т. 4. С. 55. 1094в.

² См.: Гумницкий Г. Смысл жизни, счастье, мораль. М., 1981.

II. ЛОГИЧЕСКОЕ КАНТОВЕДЕНИЕ

**Вместе с Кантом вперед к искусственному интеллекту.
Исследование оснований обработки знаний
с точки зрения «Критической методологии»**

*Марко Беттони
(Институт систем и структур, Цюрих, Швейцария)*

До сих пор считали, что всякие наши знания
должны сообразоваться с предметами...

Иммануил Кант

Обработка данных и обработка знаний

В течение последних 20 лет в компьютерных науках и в особенности в исследованиях по искусственному интеллекту (ИИ)¹ произошел сдвиг от простой обработки данных (ОД)

к значительно более сложной обработке знаний (ОЗ). Различие между двумя способами обработки можно лучше всего объяснить на примере шахматных компьютеров. Традиционный подход (ОД), по существу, заключается в том, что компьютер пытается исследовать возможно большее число положений на шахматной доске (игровых позиций). При игре средних размеров с 35 возможными ходами должны быть просмотрены $6,2 \times 10^{18}$ игровых позиций, чтобы полностью оценить последовательность из 12 ближайших ходов. При помощи более рафинированных алгоритмов (например, альфа-бета) удается это число свести к $3,1 \times 10^9$, т. е. должно быть просмотрено приблизительно 16 миллионов положений на шахматной доске в секунду, если отвести на решение о каждом ходе 3 минуты. «Deep Thought»² — в настоящее время самый мощный шахматный компьютер — может уже обрабатывать 1 миллион положений в секунду. Однако, действительно ли необходимо действовать так «неразумно» (интеллектуальная расточительность)?

Другим способом обработки, притом значительно более эффективным, является, например, способ, применяемый человеком. Экспериментальные исследования показывают, что мастер в среднем за 3 минуты исследует только около 40 положений на шахматной доске!

Если теперь попытаться понять, *каким образом* человек в той или иной области получает результаты, которые столь значительно превосходят обработку данных, то снова и снова обнаруживается, что важную роль при этом играют «знания». В случае шахмат считают, например, что в ходе игры «...планирование и образование понятий являются существенными ингредиентами...»³, и где, следовательно, планирующее и понятийное знание существенно содействует тому, чтобы избежать комбинаторного взрыва при исчерпывающем анализе и обнаружить правильный ход с помощью немногих операций мышления в более короткое время (операциональная эффективность).

Сегодня именно обработка знаний является таким способом работы компьютеров, который подражает людям не только в результатах (выбранная игровая позиция, выигранная игра), но и в функциях и механизмах (немногие операции, сложные методы). Таким образом, термин «знание» следует понимать как указание на ментальные функции и механизмы, причем я здесь понимаю под «функцией» то, что «делает возможным действие», или вместе с Кантом «условие возможности» действия.

Проблема обработки знаний

Сегодня обработка знаний потому так популярна, что она позволила совершить значительное продвижение вперед во мно-

гих областях, например, в автоматическом переводе, интерпретации данных измерений в химической технике, при диагностике болезней и технических неисправностей, при конструировании технических систем, при планировании проектов, при контроле промышленных установок и т. п.

Однако продвижение компьютеров в области, которые до этого были исключительной принадлежностью людей, не было только триумфальным шествием. Наоборот, благодаря этому удалось увидеть границы машинной обработки информации, которые ранее были неизвестны, а также четко осознать ценность ментальных функций людей.

К примеру, опыт разработки систем, основывающихся на знаниях, показывает, что главная трудность заключается в *рассчитанной на компьютер формулировке* знания, т. е. в том, что на профессиональном языке называется «проблемой представления знаний». Одновременно с внезапным появлением этого барьера исследователям в области компьютерной науки стало ясно, что человек может достигать своих лучших результатов именно благодаря своей способности к эффективнейшей и плодотворнейшей обработке и представлению знаний.

Однако, каким же образом получается, что столь оптимально функционирующую у человека организацию знания так трудно реализовать приспособленной для вычислений на компьютере форме?

Мнения ученых-компьютерщиков по поводу причин возникновения этих проблем расходятся очень сильно. Для Марвина Минского главная причина состоит в том, что искусственная обработка знаний в недостаточной степени воспроизводит ментальные процессы. Для Джона Маккарти, напротив, ОЗ еще недостаточно следует принципам математической логики. Другие же рассматривают существующие обучающие алгоритмы (нервные сети) как слишком слабые.

Радикальные критики типа Дрейфуса приходят даже к выводу, что человеческие функции обработки знаний не формализуемы, или, как считают Виноград и Флорес, опирающиеся на Хайдеггера, «нельзя сконструировать машины, которые демонстрируют или успешно моделируют разумное поведение»⁴.

Несмотря на ценные импульсы, содержащиеся в этих и других диагнозах, мне в них недостает того, что с моей точки зрения является главной причиной, а именно:

1) что в исследованиях по ИИ, касающихся «разума» (когнитивной системы), все еще недостает того, что Кант назвал «предварительной критикой способности самого чистого разума» (3, 99);

2) что фундаментом исследовательских методов в ИИ все еще является предположение, что «всякие наши знания должны сообразоваться с предметами» (3, 87).

Однако, что же я понимаю под «критикой способности ра-

зума» и чем же можно было бы заменить названный фундамент?

Чтобы объяснить это, я начну следующий раздел с некоторых цитат из «Критики чистого разума» (КЧР), которые должны на примере оригинала показать, чему ИИ может научиться у Канта. Чему я сам научился у Канта, я покажу в следующем за этим разделе, вместе с кратким представлением основных принципов новой методологии исследований в области ИИ, которую, следуя Канту, я окрестил «критической методологией».

Чему ИИ может научиться у Канта

Цитаты из КЧР, собранные в этом разделе, вкратце показывают, что «критическая методология» считает важным в Кантовой теории ментальных процессов.

В качестве приготовления прежде всего предпосылается Кантово определение знания, в соответствии с которым знание представляет собой «целое, состоящее из сопоставимых и связанных между собой представлений» (из «Дедукции чистых рассудочных понятий» (3, 700)).

В следующей методологически важной цитате Кант обосновывает при помощи очень простого и практического («коперниканского») рассуждения новое, и сегодня все еще революционное направление, которое он выбрал для своих исследований: «До сих пор считали, что всякие наши знания должны соотнобразоваться с предметами. При этом, однако, кончались неудачей все попытки через понятия что-то априорно установить относительно предметов, что расширяло бы наше знание о них. Поэтому следовало бы попытаться выяснить, не разрешим ли мы задачи метафизики более успешно, если будем исходить из предположения, что предметы должны соотнобразоваться с нашим познанием,— а это лучше согласуется с требованием возможности априорного знания о них, которое должно установить нечто о предметах раньше, чем они нам даны» (3, 87, курсив мой.— М. Б.).

Главный вопрос здесь заключается в том, каким образом предметы могут соотнобразоваться с «нашим познанием», а именно с ментальными процессами нашей обработки знаний. Решение, предложенное Кантом в КЧР, заключается в том, что предметы конституированы ментально. Принимается во внимание, что мы в этом случае имеем два класса предметов, а именно, с одной стороны, многообразное физическое ситуации (неопределенное), а с другой стороны, объект, который мы конституируем ментально (определенное) и который также может быть назван «объектом-для-меня». Чтобы конституировать такие объекты-для-меня, когнитивная система (английское «mind») использует операции (действия) синтеза, который должен обнаруживаться в процессе обработки знания на трех

различных ступенях и который необходим, чтобы обосновать единство (интеграцию) сознания. Кант отводит синтезу высочайшее место, он возвышается даже до центрального понятия всей его «критической философии»⁵. Я придерживаюсь того мнения, что если бы нам удалось должным образом реконструировать подходящую модель кантовского понимания «синтеза», то мы были бы в состоянии тем самым совершить революционное продвижение в «когнитивной» науке.

Первое определение синтеза во втором издании КЧР имеет следующий вид: «Однако спонтанность нашего мышления требует, чтобы это многообразное прежде всего было каким-то образом просмотрено, воспринято и связано для получения из него знания. Такое действие я называю синтезом» (3, 173). «Синтез многообразного (будь оно дано эмпирически или а priori), порождает прежде всего знание...; ...именно синтез есть то, что, собственно, составляет из элементов знание и объединяет их в определенное содержание. Поэтому синтез есть первое, на что мы должны обратить внимание, если хотим судить о происхождении наших знаний» (3, 173). В этом определении мы можем обнаружить четыре важных установки. Во-первых, операндой, на которую действует синтез, является многообразное, которое может быть дано или а priori, или эмпирически. Во-вторых, частная функция, из которой получается главная функция синтеза, есть не что иное, как связь (английское «connection»). В-третьих, процесс обработки знаний, а именно возникновение каждой добавочной единицы знания («некоторого знания», «некоторого содержания») начинается с операции синтеза. В-четвертых, сначала следует промоделировать способность синтеза (а не анализа), если стремиться понять наш способ обработки знаний.

Частичная функция синтеза в Кантовой модели ментальной обработки работает на различных ступенях сложной системы обработки знаний. Одной из важнейших ступеней является та, которая имеет своим результатом «объект-для-меня», а именно конституируемый объект. Однако объекты в когнитивной системе порождаются не как материальные сущности («не в смысле существования» (3, 187), и создаются не как ментальные копии реальности: они конституированы (сделаны возможными, а priori установлены (3, 187) как ментальные операнды (= «представления») посредством синтеза многообразного.

Понятие «связи» играет центральную роль в Кантовой модели. Он вводит это понятие в «Трансцендентальной дедукции чистых рассудочных понятий», в части, которая образует ядро и первого, и второго издания и которая во втором издании была полностью переработана.

В следующей цитате Кант подчеркивает активное, продуктивное существо ментальной функции, которую он называет связью (в противоположность пассивному, репродуктивному

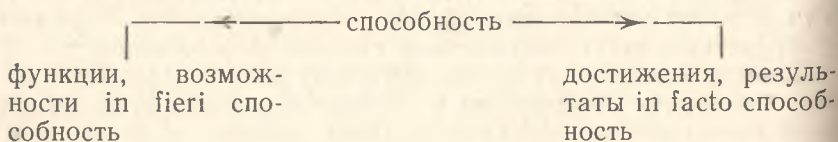
существо других функций), благодаря чему должно быть ясно, что предметы в когнитивных системах не могут быть копиями физических предметов, но являются *ментально детерминированными* изображениями ментально построенного единого мира обработки знаний: «Однако не предмет заключает в себе связь, которую можно заимствовать из него путем восприятия, только благодаря чему она может быть усмотрена рассудком, а сама связь есть функция рассудка, а сам рассудок есть не что иное как способность а priori связывать и подводить многообразное (содержание) данных представлений под единство апперцепции. Этот принцип есть высшее основание во всем человеческом знании» (3, 193).

Теперь мы можем обобщить Кантово понимание синтеза следующим образом: операции синтеза порождают связь и его результатом является предмет как ментально детерминированный объект, объект-для-меня (объектность обработки знаний). Именно этот «синтез объектов-для-меня» предоставляет материал для анализа. Лишь после того, как дан этот материал, оказываются возможными ментальные процессы типа «объективного отличительного признака» и «объективной классификации» (которые господствуют сегодня в моделях ИИ). Это соответствует и Кантову основному положению, согласно которому «там, где рассудок ничего раньше не связал, ему нечего и разлагать» (3, 190).

«Критическая методология»

Прежде чем приступить к краткому изложению того, каким образом могла бы выглядеть новая методология исследований в области ИИ, я должен сначала ввести еще три вспомогательные дистинкции.

Первая дистинкция заключается в различении двух точек зрения на «способность» (*Vermögen*) (фигура 1): с одной стороны, «*in facto* * способности», а с другой стороны, «*in fieri* ** способности». С первой точки зрения (*in facto*), способность имеет значение *достижений* в смысле «достигнутого» или результата. Со второй точки зрения, оно обозначает *функции* в смысле «возможностей», возможных действий.



Фигура 1. Значения «способности» с двух точек зрения. Теперь я проведу различие (вторая дистинкция)

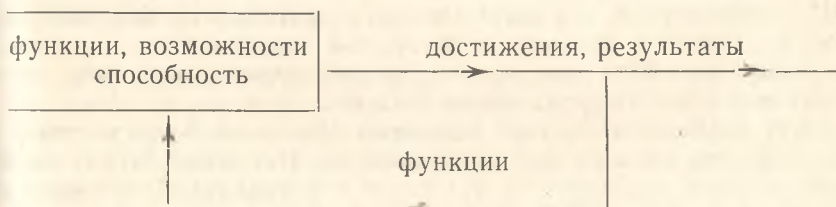
* На деле, фактически (лат.) — Прим. перев.

** В становлении (лат.) — Прим. перев.

между тремя видами в рамках и способностей-результатов и способностей-функций (фиг. 2). Первый — это символы, как результаты обработки символов. Второй — это объекты, как результаты обработки объектов, и третий — это знание, т. е. содержание мышления, как результат соединения обработки объектов и символов в когнитивной системе.



Фигура 2. Структурирование «знания» на две компоненты. Третью дистинкцию я получаю, вводя между результатами и функциями условие, согласно которому сами результаты функций снова могут быть функциями («mens agitat mentem»)*, что может быть описано с помощью понятий «аутопоэзиса» и «самоорганизации».



Фигура 3. Функции как результаты функций. Критическая методология (сокращенно КМ) начинается с того, что Кантова установка принимается всерьез, и в соответствии с ней объясняется, почему «кончались неудачей все попытки через понятия что-то априорно установить относительно предметов...» (3, 87).

С точки зрения критической методологии, эта установка верна и для состояния исследований по ИИ, т. е. в ИИ также ничего нельзя «установить» *через понятия* относительно основной проблемы обработки знаний. Поэтому необходимы принципиально новые методы исследования основных проблем ИИ вообще и «когнитивной науки» в особенности. Главной частью этой методологии является «Критика способности разума».

Под этим я понимаю такое исследование *результатов* (достижений, английское «performances») обработки знаний, которое ставит под сомнение эти ментальные достижения (отсюда предикат «критический») до тех пор, пока не будет разработана система, теория, модель для *функций*, т. е. для «условий возможности» этих достижений.

В силу центральной роли понятий для ментальной обработки знаний КМ предлагает, чтобы понятия (как ментальные ре-

* Ум двигает умом (лат.) — Прим. перев.

зультаты) до тех пор ставились под сомнение, пока не будет разработана система для функций обработки понятий. Кант разработал достаточно глубокую систему такого рода в «Трансцендентальном учении о началах». Основные принципы современного ИИ ясно свидетельствуют, что такая постановка под сомнение не совершается в сегодняшнем ИИ и что такой ИИ не способен разработать систему такого рода. Один из таких принципов, на которых для всех основывался классический ИИ, гласит, что «физическая система символов необходима и достаточна для разумных действий»⁶. Другой принцип, на который опираются исследователи «коннекционизма» (РОР, нервные сети), гласит, что «интеллект возникает из взаимодействия большого числа простых единиц обработки»⁷. Третий принцип является общим для всех направлений ИИ и выразительно гласит, что ИИ «не связан с каким-либо определенным способом получения результатов»⁸, а четвертый принцип, который скорее неявно, чем явно, обнаруживается в исследованиях по ИИ, проявляется в абсолютистских притязаниях формальной логики, которая целиком и полностью господствует на сцене, не оставляя места для других теорий логики, например, для Кантовой «трансцендентальной логики».

КМ требует, чтобы эти основные принципы были поставлены под сомнение и модифицированы. Исходный пункт этой методологии порывает с эмпиристской традицией, которая в исследованиях по ИИ достигла своей высшей точки,— но также и своего величайшего кризиса,— и состоит в сдвиге, аналогичном Кантову «коперниканскому перевороту», призывая исследователя отважиться на необычное предприятие. Следовательно, исходный пункт критической методологии в свете Кантовой формулировки в КЧР гласит: «Следовало бы попытаться выяснить, разрешим ли мы более успешно задачи ИИ, если будем исходить из предположения, что предметы должны соотноситься с нашими ментальными процессами».

Исходя из этого, в КМ выдвигается следующая рабочая гипотеза: *логическая организация* сигналов, обрабатываемых когнитивной системой (включая входные сигналы), является целиком и полностью нашим ментальным актом, следовательно, она принадлежит нашему ментальному миру и не существует в физическом мире. Кант выражает эту гипотезу там, где он говорит: «Поэтому следовало бы попытаться выяснить, не разрешим ли мы задачи... более успешно, если будем исходить из предположения, что предметы должны соотносываться с нашим познанием...» (3, 87).

Следующие замечания должны дополнить и разъяснить этот центральный пункт:

1) Каждый входной сигнал, который становится частью моего когнитивного акта, тем самым предполагает *ментальную операцию синтеза*, которая встраивает его в некоторую логиче-

скую организацию, придает ему «внутреннюю форму» (поэтому слово «информация» так хорошо подходит к ментальным процессам).

2) Предметы без логической организации неопределенны (т. е. могут не существовать в КС), а логическая организация без предметов пуста (т. е. существует, не принося познания).

3) Логическая организация предметов (как операнд когнитивной системы), которая в Кантовом учении зависит от «синтеза», «связи», «синтетического единства апперцепции» и т. п., строится только самой КС; она никогда не «импортируется» с сигналами.

4) Только логическая организация создает из многообразия физически организованных сигналов то, что мы воспринимаем как предмет (синтетически произведенный предмет).

5) Мы не можем воспринимать никаких сигналов, не организуя их логически при помощи функции синтеза. Но если мы организуем их логически, то мы *определяем* их, они начинают существовать в нашей голове и *сообразуются* с нашими процессами.

6) Вопреки наивному представлению «здорового рассудка» и многих исследователей ИИ, вход, который содержит когнитивная система, связан исключительно с физической организацией и логически неопределен, т. е. не имеет какого-либо *логического* содержания.

7) Физическая организация определяет объект только физически «в порядке существования», но не достаточна для того, чтобы в когнитивной системе получить объект (познавать), вопреки принципам коннекционизма.

8) *Объект* существует в когнитивной системе только в силу того, что посредством наших ментальных операций мы соединяем физическую организацию сигналов (многообразное) при помощи логической организации в единое целое и манипулируем ими, т. е. к физической определенности добавляется логическая определенность.

9) Если объект определяется таким образом, то тем самым облегчается и становится более эффективной его дальнейшая обработка, которая уже связана с некоторым символом. *Символ*, таким образом, есть не что иное, как указатель или заместитель построенного нами объекта.

10) Объект и символ *вместе* составляют наименьшую единицу знания, следовательно, символы без объектов (классический ИИ) или физические объекты без логической организации (коннекционизм) не могут дать обработки знаний.

11) Целостность, единство, связность логической и физической организации представляет собой условие возникновения каждого предмета как операнды когнитивной системы, и там, где наше познание связано исключительно с переработкой таких операнд, — также составляет условие всякого знания.

Моя попытка *продвинуться вместе с Кантом по направлению к ИИ* началась с кибернетической интерпретации КЧР¹⁰. Критическая методология представляет собой продолжение такой попытки, но пока еще находится в самом начале своего развития. Ближайший шаг, который предстоит сделать при помощи исследовательского проекта, в рамках которого я собираюсь заняться разработкой компьютерного языка для обработки знаний¹¹, одновременно послужит дополнительной проверкой основных принципов моего подхода.

¹ В статье использованы следующие сокращения: ИИ — искусственный интеллект (Künstliche Intelligenz, Artificial Intelligence); ПЗ — представление знаний (Wissens-Darstellung, Knowledge Representation); ОЗ — обработка знаний (Wissens-Verarbeitung, Knowledge Processing); КМ — критическая методология (Kritische Methodologie, Critical Methodology); КС — когнитивная система (Kognitives System, mind); КЧР — «Критика чистого разума».

² Reddy R. Foundations and Grand Challenges of Artificial Intelligence//AI-Magazine. 1988. Vol. 9. N 1. P. 16.

³ Levy D. Master Level and Beyond//Levy D. Computer Games. N. Y.: Springer, 1988. Vol. 1. P. 206.

⁴ Winograd T., Flores F. Understanding Computers and Cognition. Norwood: Ablex. 1986. P. 11.

⁵ Wolff R. Kant's Theory of Mental Activity. Gloucester: Peter Smith, 1973. P. 67.

⁶ Reddy R. Op. cit. P. 13.

⁷ Rumelhart D., McClelland J. Parallel Distributed Processing. Boston: MIT Press, 1986. P. IX.

⁸ Charniak E., McDermot D. Introduction to Artificial Intelligence. Reading: Addison-Wesley, 1985.

⁹ Genesereth M., Nilsson N. Logical Foundations of Artificial Intelligence. Los Altos: Morgan Kaufman, 1987.

¹⁰ Bettoni M. Cybernetics applied to Kant's Architecture of Mind//7-th International Kant-Congress. Mainz, 1990.

¹¹ Bettoni M. An Abstract Language for Knowledge Representation Based on Patterns of Mental Quanta//8-th International Congress of Cybernetics and Systems. N. Y., 1990.

Перевод с нем. В. Н. Брюшкина

Кант и «искусственный интеллект»: трансцендентальный анализ моделей мира

*В. Н. Брюшкин
(Калининградский государственный университет)*

По проблеме «Кант и «искусственный интеллект» мною были ранее опубликованы статьи в 14-м и 15-м выпусках «Кантовского сборника». Настоящая статья продолжает исследование этой проблемы. В предшествовавшей публикации¹ я рассмотрел способ построения моделей мира, который можно извлечь из Кантовой «Аналитики понятий». Однако в ней мною была проявлена некоторая поспешность, именно в той части,