Возможна ли алгоритмическая реализация интеллекта и какие ее ограничения

Вопросы для обсуждения

- Верно ли мнение Х. Дрейфуса («Алхимия и искусственный интеллект», «Чего не могут вычислительные машины»)?
- О чем говорит парадокс Дж. Сёрла «Китайская комната»?
- Насколько прав Р. Пенроуз («Новый ум короля», «Тени разума»)?
- Накладывает ли теорема Гёделя какие-то ограничения на возможности алгоритмического ИИ?
- Нужна ли новая математика для создания ИИ?

ALCHEMY AND ARTIFICIAL INTELLIGENCE

Hubert L. Dreyfus

December 1965

Early successes in programming digital computers to exhibit simple forms of intelligent behavior, coupled with the belief that intelligent activities differ only in their degree of complexity, have led to the conviction that the information processing underlying any cognitive performance can be formulated in a program and thus simulated on a digital computer. Attempts to simulate cognitive processes on computers have, however, run into greater difficulties than anticipated.

An examination of these difficulties reveals that the attempt to analyze intelligent behavior in digital computer language systematically excludes three fundamental human forms of information processing (fringe consciousness, essence/accident discrimination, and ambiguity tolerance). Moreover, there are four distinct types of intelligent activity, only two of which do not presuppose these human forms of information processing and can therefore be programmed. Significant developments in artificial intelligence in the remaining two areas must await computers of an entirely different sort, of which the only existing prototype is the little-understood human brain.

CLASSIFICATION OF INTELLIGENT ACTIVITIES

I. Associationistic	II. Non-formal	III. Simple Formal	IV. Complex Formal
Characteristics of Activity			
Irrelevance of meaning and context.	Dependent on meaning and context, which are not explicit.	Meanings completely explicit and context-independent.	In principle, same as III; in practice, internally context-dependent, independent of external context.
Learned by repetition.	Learned by perspicuous examples.	Learned by rule (exception: checkers).	Learned by rule and practice.
Kinds of Program			
Decision tree, list search.	None.	Algorithm or limit on growth of search tree.	Search-pruning heuristics.

Without the associationist assumption to fall back on, what encouragement can workers in cognitive simulation and artificial intelligence draw from the argument that, even though the brain does not process information in discrete operations, we can simulate the brain on a digital computer and thus by discrete operations produce the same results the brain produces?

Similarly, one must delimit what can count as information processing in a computer. A digital computer solving the equations describing an analogue information-processing device and this simulating its <u>function</u> is not thereby simulating its <u>information processing</u>. It is not processing the information which is processed by the simulated analogue, but entirely different information concerning the physical or chemical properties of the analogue. Thus the strong claim that every processable form of information can be processed by a digital computer is misleading. One can only show that, for any given type of information, a digital computer can in principle be programmed to simulate a device which can process that information. This does not support Feigenbaum's assertion that human information processes are perfectly explicable, and therefore fails to show that "digital computers being general information-processing devices, they can be programmed to carry out any and all information processes."

Выводы

- Дискретность vs аналоговость
- Неформальность
- Сознание и понимание
- Модель отлична от феномена
- Требуются принципиально иные компьютеры

Searle «Minds, Brains, and Programs»

Аксиома 1. Компьютерные программы — это формальные (синтаксические) объекты.

Аксиома 2. Человеческий разум оперирует смысловым содержанием (семантикой).

Аксиома 3. Синтаксис сам по себе не составляет семантику и его недостаточно для существования семантики.

Следствия

Следствие 1: Программы не являются сущностью разума и их наличия недостаточно для наличия разума.

Следствие 2: Тот способ, посредством которого человеческий мозг на самом деле порождает ментальные явления, не может сводиться лишь к выполнению компьютерной программы.

Следствие 3: То, что порождает разум, должно обладать по крайней мере причинно-следственными свойствами, эквивалентными соответствующим свойствам мозга.

Комментарии Сёрла

1. Во-первых, я не пытался доказывать, что «компьютер не может мыслить». Поскольку все, что поддается моделированию вычислениями, может быть описано как компьютер, и поскольку наш мозг на некоторых уровнях поддается моделированию, то отсюда тривиально следует, что наш мозг — это компьютер, и он, разумеется, способен мыслить. Однако из того факта, что систему можно моделировать посредством манипулирования символами и что она способна мыслить, вовсе не следует, что способность к мышлению эквивалентна способности к манипулированию формальными символами.
2. Во-вторых, я не пытался доказывать, что только системы биологической природы, подобные нашему мозгу, способны мыслить. В настоящее время это единственные известные нам системы, обладающие такой способностью, однако мы можем встретить во Вселенной и другие способные к осознанным мыслям системы, а может быть, мы даже сумеем искусственно создать мыслящие системы. Я считают этот вопрос открытым для споров.

Комментарии Сёрла

3. В-третьих, утверждение «сильного» ИИ заключается не в том, что компьютеры с правильными программами могут мыслить, что они могут обладать какими-то неведомыми доселе психологическими свойствами; скорее, оно состоит в том, что компьютеры просто должны мыслить, поскольку их работа — это и есть не что иное, как мышление.

4. В-четвертых, я попытался опровергнуть сильный ИИ, определенный именно таким образом. Я пытался доказать, что мышление не сводится к программам, потому что программа лишь манипулирует формальными символами — а, как нам известно, самого по себе манипулирования символами недостаточно, чтобы гарантировать наличие смысла. Это тот принцип, на котором основано рассуждение о китайской комнате.

Пенроуз «Тени разума»

«…здесь я решительно выступаю против широко распространенного мнения, согласно которому нашу сознательную мыслительную деятельность - во всех ее разнообразных проявлениях - можно, в принципе, адекватно описать в рамках тех или иных вычислительных моделей... предполагает подлинный поиск (...) инструментов, позволяющих описываемому в научных терминах мозгу применять для осуществления требуемой невычислительной деятельности тонкие и по большей части нам пока не известные физические принципы.

Еще менее способны вычисления, взятые сами по себе, обусловить какое бы то ни было сознательное ощущение или желание.

Иначе говоря, я полагаю, что упомянутая квантовая активность должна быть неким невычислимым образом связана с поддающимся вычислению процессом »

Точки зрения

A. Всякое мышление есть вычисление; в частности, ощущение осмысленного осознания есть не что иное, как результат выполнения соответствующего вычисления.
B. Осознание представляет собой характерное проявление физической активности мозга; хотя любую физическую активность можно моделировать посредством той или иной совокупности вычислений, численное моделирование как таковое не способно вызвать осознание.
C. Осознание является результатом соответствующей физической активности мозга, однако эту физическую активность невозможно должным образом смоделировать вычислительными средствами.
D. Осознание невозможно объяснить в физических, математических и вообще научных терминах.
Точка зрения D... сродни религиозной доктрине.

В: Научный здравый смысл?

«Как отмечает философ Джон Серл, вычислительную модель физического процесса никоим образом не следует отождествлять с самим процессом, происходящим в действительности. (Компьютерная модель, например, урагана - это совсем не то же самое, что и реальный ураган!) Согласно взгляду В, наличие или отсутствие сознания очень сильно зависит от того, какой именно физический объект "осуществляет мышление" и какие физические действия он при этом совершает.»

«Допускает ли современная физика существования процессов, которые принципиально невозможно смоделировать на компьютере? Если мы надеемся получить на этот вопрос математически строгий ответ, то нас ждет разочарование... Однако сам я убежден в том, что подобные невычислимые процессы следует искать за пределами тех областей физики, которые описываются известными на настоящий момент физическими законами.»

«все нормальные системы, к которым применим термин "хаотические", следует относить к категории систем, которые я называю "вычислительными". »

Непрерывность

«В случае аналоговых систем необходимо учитывать одно формальное обстоятельство: стандартные понятия вычисления и вычислимости применимы, строго говоря, только к дискретным системам (...), но не к непрерывным, таким, например, как расстояния или электрические потенциалы, с которыми имеет дело традиционная классическая физика... при компьютерном моделировании физических систем вообще стандартной процедурой является аппроксимация всех рассматриваемых непрерывных параметров в дискретной форме.

Существуют, однако, и другие подходы к проблемам вычислений... непрерывные системы рассматриваются как самостоятельные математические структуры со своим собственным понятием "вычислимости" - понятием, обобщающим идею вычислимости по Тьюрингу с дискретных величин на непрерывные.

В соответствии с сильной версией С, невычислимость сознательной деятельности мозга может быть исчерпывающе объяснена в рамках некоторой невычислительной физической теории (пока еще не открытой), следствия из которой, собственно, и обуславливают упомянутую деятельность. Хотя второй вариант может показаться несколько надуманным, альтернатива (для сторонников С) и в самом деле состоит в отыскании для какого-либо непрерывного процесса в рамках известных физических законов такой роли, которую невозможно было бы адекватно воспроизвести посредством каких угодно вычислений.»

Роль окружения

«А что нам известно о роли окружения?.. Возможно, именно это уникальное личное окружение и даст каждому из нас ту особенную последовательность входных данных, которая неподвластна вычислению?..
А может быть, численное моделирование пусть даже всего лишь правдоподобного окружения невозможно в принципе. Быть может, в окружающем физическом мире все же есть нечто такое, что на самом деле неподвластно численному моделированию. Возможно, некоторые сторонники А или В уже вознамерились приписать все не поддающиеся, на первый взгляд, вычислению проявления человеческого поведения невычислимости внешнего окружения... как только мы признаем, что физическое поведение допускает где-то что-то такое, что невозможно моделировать вычислительными методами, мы тем самым тут же лишаемся главного, по всей видимости, основания сомневаться в правдоподобии... С.»

Теорема Гёделя

«Так что же такое теорема Гёделя? В 1930 году на конференции в Кёнигсберге блестящий молодой математик Курт Гёдель произвел немалое впечатление на ведущих математиков и логиков со всего мира, представив их вниманию теорему, которая впоследствии получила его имя. Ее довольно быстро признали в качестве фундаментального вклада в основы математики - быть может, наиболее фундаментального из всех возможных, - я же, в свою очередь, утверждаю, что своей теоремой Гёдель также положил начало важнейшему этапу развития философии разума.

... нельзя создать такую формальную систему логически обоснованных математических правил доказательства, которой было бы достаточно, хотя бы в принципе, для доказательства всех истинных теорем элементарной арифметики.»

«Для установления математической истины математики не применяют заведомо обоснованные алгоритмы.»

«Они показали, что определенные математические утверждения (так называемые континуум-гипотеза Кантора и аксиома выбора) никак не зависят от теоретико-множественных аксиом системы Цермело-Френкеля... Согласно же континуум-гипотезе Кантора, количество подмножеств натуральных чисел - равное количеству вещественных чисел - представляет собой вторую по величине бесконечность после множества собственно натуральных чисел.»

«Многие ошибочно полагают (), что из теоремы Гёделя следует существование множества различных арифметик, каждая из которых в равной степени обоснованна. Соответственно, та частная арифметика, которую мы, возможно, по чистой случайности избрали для своих нужд, определяется просто какой-то произвольно взятой формальной системой. В действительности же теорема Гёделя показывает, что ни одна из этих формальных систем (будучи непротиворечивой) не может быть полной... Гёдель доказал не то, что математика () - это произвольные поиски, направление которых определяется прихотью Человека; он доказал, что математика - это нечто абсолютное, и в ней мы должны не изобретать, но открывать.»

Выводы

- Алгоритмическая неразрешимость
- Сознание и понимание
- Модель отлична от феномена
- Понимание – физический процесс
- Требуются принципиально иные компьютеры

Нужна ли некорректная математика для решения некорректных задач?