

Хьюберт Дрейфус об ИИ и возможностях его создания

Введение

Хьюберт Дрейфус - американский философ, представитель феноменологии. В своих книгах *Alchemy and AI*, *What computers can't do*, *Mind Over Machine* Дрейфус выдвигает ряд аргументов, которые по его мнению делают невозможным создание искусственного разума. Дрейфус считает, что человеческие рассуждения и интеллектуальное поведение зависят в первую очередь от бессознательных инстинктов в противовес сознательной манипуляции с символической информацией. По Дрейфусу эти бессознательные умения, которыми обладает человек, никогда не могут быть воплощены в машине с использованием формальных правил.

Исторический контекст

Прежде чем перейти к подробному обсуждению идей Дрейфуса и тому, какое влияние они оказали на науки об искусственном интеллекте, стоит кратко обозреть исторический контекст, имевший место на момент написания его книг.

В начале книги "Что не могут компьютеры" Дрейфус рассматривает мысли философов и ученых прошлого, которые явились предпосылками к идее о том, что машина может мыслить. Он прослеживает этот процесс от

- Платона (ты знаешь только то, что можешь выразить в виде правил, которые всем понятны. Живописцы не знают ничего, они божественно одарены),
- через Гоббса (рассуждение суть суммирование частей),
- Лейбница (я ищу синтаксические кванты, путем сложения которых можно производить любые рассуждения),
- Буля и Бэббиджа до Тьюринга (если машину в разговоре не отличить от человека - то мы будем называть её мыслящей)

Книги Дрейфуса написаны в промежутке от середины шестидесятых до конца семидесятых годов. К тому моменту имелись некоторые успехи в области создания нейронных сетей, машинного перевода, начали разрабатывать игровые и в первую очередь шахматные программы, возникло несколько программ, имитирующих человеческие рассуждения. Примерами последних являлись GPS (General Problem solver, 1957), успешно решавший алгоритмические задачи, программа Вайценбаума ЭЛИЗА - компьютерный собеседник, сумевший одурачить психотерапевтов.

После первых ошеломительных успехов исследователи ИИ начали делать смелые, порой даже рекламные заявления. Приведем примеры:

Смелые заявления сторонников ИИ

1958, Саймон, Ньюэлл автор GPS:

1. Не пройдет и десяти лет, как цифровая вычислительная машина станет чемпионом мира по шахматам, если не будут введены правила, не допускающие ее к соревнованиям.
2. Не пройдет и 10 лет, как вычислительная машина найдет и докажет важную и до сих пор неизвестную математическую теорему.
3. Не пройдет и десяти лет, как большинство психологических теорий примет форму программ для вычислительных машин или качественных утверждений о тех или иных характеристиках машинных программ

1965, Саймон, автор GPS

1. В ближайшие двадцать лет машины будут способны делать любую работу, которую сейчас делают люди

1967, Marvin Minsky

уже следующее поколение в основном решит проблему создания искусственного интеллекта.

1970, Marvin Minsky (in Life Magazine)

В течение трех-восьми лет мы будем иметь машину, обладающую интеллектом среднего человеческого существа

На что нападает Дрейфус?

Специалист в области новейшей европейской философии, феноменолог по убеждениям, Дрейфус усомнился в том, что какой-либо из этих позитивных прогнозов окажется правдивым. Давайте рассмотрим, против чего он борется, взятое из книги "Чего не могут компьютеры":

" ИИ представляет собой попытку моделирования разумного поведения человека с помощью таких методов программирования, которые не имеют или почти не имеют сходства с мыслительными процессами человека. "

Допущения, на которых базируется программа создания иск.интеллекта

С изумлением наблюдая оптимизм и неуемную энергию, брызжущую из исследователей ИИ, Дрейфус задумывается, почему же они считают свою цель заведомо достижимой? Он приходит к четырем допущениям, которые в той или иной мере разделяются приверженцами AI:

1. **Биологическое допущение:** на некотором уровне - обычно полагают, что на уровне нейронов, -операции по переработке информации носят дискретный характер и

- происходят на основе некоторого биологического эквивалента переключательных схем,
2. **Психологическое допущение:** мышление можно рассматривать как переработку информации, заданной в бинарном (двоичном) коде, причем переработка происходит в соответствии с некоторыми формальными правилами. Таким образом, в психологии вычислительная машина служит в качестве модели рассудка, каким его представляли эмпирики, такие, например, как Д.Юм (в этом случае информационным “битами” соответствуют атомарные впечатления), или идеалисты вроде И.Канта {в этом случае программа реализует правила мыслительного процесса) .
 3. **Эпистемологическое допущение:** все знания могут быть формализованы, то есть все, что может быть понято, может быть выражено в терминах логических отношений, точнее, в терминах булевых функций - логического исчисления, задающего правила обращения с информацией, заданной в двоичном коде
 4. Наконец, поскольку вся информация, которая вводится в машину, должна быть представлена в двоичной форме - в битах,- машинная модель мышления предполагает, что все сведения о мире, все, что составляет основу разумного поведения, должно в принципе допускать анализ в терминах множества элементов, безразличных к ситуациям. Таково **онтологическое допущение:** все, происходящее в мире, можно представить в виде множества фактов, каждый из которых логически не зависит от остальных.

Опровержение допущений

Биологическое

Биологическое допущение опровергается чисто эмпирическим путем. Это одно из самых сильных и наглых допущений и на момент написания работы с Дрейфусом никто не спорил о том, что биологическое допущение оправдано.

Из того факта, что нервные импульсы между нейронами построены по принципу “все или ничего”, вовсе не следует, что имеет место цифровой процесс в любой его форме. Различие между цифровым и аналоговым вычислением - это логическое различие, которое не зависит ни от конструкции системы, ни от вида используемых в ней электрических импульсов.

Существенное различие между цифровым и аналоговым способами переработки информации состоит в том, что при цифровом способе отдельный элемент представляет символ некоторого описательного языка, то есть несет конкретный квант информации, тогда как в устройствах, работающих по аналоговому принципу, подлежащая обработке информация представляется непрерывными физическими величинами. Мозг, оперирующий потоками импульсов, можно считать цифровым вычислительным устройством только в том случае, если каждый импульс окажется связанным с определенным шагом в процессе переработки информации; если же, напротив, обнаружится, что минимальной “единицей” в модели информационного процесса должна быть частота импульсации, то мозг работает как аналоговое устройство.

Психологическое

Психологическое допущение, говорящее о том, что **разум может рассматриваться как устройство, обрабатывающее символьную информацию по формальным правилам; разум работает на тех же принципах, что и цифровая вычислительная машина.**

Дрейфус нападает на это допущение довольно агрессивно - ведь если мы его примем, то психологические высказывания и исследования будут вестись в терминах программ, что по Дрейфусу может повлиять на податливую человеческую психику. "Нам грозит не пришествие сверхразумных машин, а появление неполноценно мыслящих человеческих существ" - замечает Хьюберт.

Дрейфус отвергает психологическое допущение, показывая, что многое из того, что большая часть наших о мире представлена не в виде формальных правил, а в виде комплекса взглядов, точек зрения и контекстов, позволяющих человеку работать с различными интерпретациями знаний. Даже когда мы манипулируем формальными символами, мы используем их, держа в уме большой массив априорных, фоновых знаний, здравого смысла. Без этого массива манипуляции с формальными символами перестают что-либо значить.

Рассмотрим одну из иллюстраций, приводимую Дрейфусом. Он оппонирует Джерри Фодору в следующем утверждении:

"Пусть имеется яма и человек на глаз способен определить её глубину. Если причиной определения глубины является текстурный градиент, то мы можем описывать физические процессы, происходящие в нервной системе именно в этих терминах: "вычисление информации о текстурных градиентах"

Дрейфус замечает, что до тех пор пока непонятно, как именно МОЗГ определяет глубину, мы не можем утверждать, что он вычисляет текстурный градиент. Это будет так же абсурдно, как утверждение о том, что обращаясь вокруг солнца планеты решают диф.уравнения. **Из того факта, что все психохимические процессы могут быть формализованы и рассчитаны в дискретной форме вовсе не следует, что какие-либо дискретные процессы имели место в действительности.** А для психологии нам недостаточно просто смоделировать человеческий разум с тем, чтобы не было различий в результатах рассуждений - нет, нам нужно более сильное моделирование, учитывающее способ вычислений, производимый разумом.

Эпистемологическое

Перейдем теперь к эпистемологическому допущению. Episteme - знание, а допущение говорит о том, что

Тезис А: **все знания формализуемы и могут быть выражены в битовой форме.**

В такой формулировке это допущение формулируется Дрейфусом во введении к второй части книги. В главе "эпистемологическое допущение" Дрейфус приводит более слабую формулировку тезиса, который и опровергает на тридцати страницах. Приведем эту новую формулировку:

Тезис Б: **"Хотя действия человека нельзя объяснить на основе предположения, пользуясь лишь эвристическими правилами и последовательностями операций, разумное поведение может быть формализовано в терминах такого рода правил и благодаря этому воспроизведено на машине. "**

Вглядевшись в эти два тезиса мы видим, что Б является частью А. Проследим ход мыслей философа при опровержении тезиса Б.

Две части эпистемологического допущения

Во первых Дрейфус делит это допущение на две части:

1. всякое произвольное поведение может быть формализовано
2. для воспроизведения интересующего нас поведения может быть использован формальный метод.

Формализация произвольного поведения

Сначала рассмотрим первую часть: формализация произвольного поведения. Мы можем делать это физически, моделируя мозг, как физическую систему, а также лингвистически - формализуя лингвистику.

Физическое моделирование потребует слишком больших ресурсов и временных затрат - Дрейфус ссылается на то, что возможность этого моделирования противоречит законам физики и теории информации.

Лингвистического моделирования недостаточно для достижения целей ИИ. Хомский и представители трансформационной лингвистики доказали возможность формализации грамматики, а для ИИ нужны способы формализации семантики и прагматики языка. Иллюстрация: выражение "мысль еще не дошла до бумаги".

Воспроизведение поведения формальными методами

Доказательство второй части аналогично первой - в отсутствии формализации того, как вести себя в контексте, мы не можем моделировать поведение формальными методами.

Онтологическое

Онтологическое допущение (ontos - бытие), состоящее в том, что мир может быть проанализирован в терминах четко определенных атомарных фактов, плавно вытекает из эпистемологического допущения.

Дрейфус показывает, что онтологическое допущение формировалось очень долгое время и стало неназванной аксиомой в современной философской традиции. В начале упоминались Платон, Лейбниц, Юм, Рассел с его логическим атомизмом. Наиболее явно это выражено "Логико-философском трактате" Л. Витгенштейна; в этом труде мир определяется в терминах множества атомарных фактов, которые могут быть выражены логически независимыми предложениями.

Феноменологи 20 века Мерло-Понти и Хайдеггер выступают против аксиоматизации онтологического допущения, считая его только одной из гипотез.

Аргументация Дрейфуса довольно изощрена и мы не будем приводить её в этом докладе. Заметим только, что эмпирические кванты сознания, контекстно-независимые кусочки мира пока найдены не были, что позволяет усомниться в онтологическом допущении.

Заключение

Возможно, Дрейфус был слишком категоричен в своих суждениях; успехи последних лет в области искусственного интеллекта опровергают излишне пессимистические заявления Дрейфуса и других критиков. Однако хотелось бы отметить, что столь конструктивная критика оснований и допущений, делаемых учеными, заставила многих исследователей задуматься и приступить к осмыслению новых, более перспективных путей развития.

Философское осмысление вопросов ИИ и впрямь поспособствовало выработке более трезвого взгляда на то, чего можно и чего нельзя ожидать от такого рода устройств.

From:

<http://wiki.osll.ru/> - **Open Source & Linux Lab**

Permanent link:

http://wiki.osll.ru/doku.php/etc:users:yuri_v_katkov:dreyfus

Last update: **2016/08/08 20:53**

