

Я не знаю другой такой области когнитивной психологии, где происходили бы столь ожесточенные споры о моделировании человеческого мышления машинами: о "жестком" направлении ИИ. Одну сторону этого спора представляют те ревнители ИИ, которые не только верят, что машины способны точно копировать человеческое познание, но и считают, что наиболее сложные интеллектуальные процессы могут выполняться только машинами. Это надо понимать так, что компьютеры должны непосредственно участвовать в повседневном принятии решений людьми. С другой стороны, находятся те, кто полагает ИИ интеллектуально извращенным понятием и считает, что люди, верящие в "мыслящие машины", — это материалистические идолопоклонники. Они полагают, что человеческое мышление — это чисто человеческий процесс; наверно, его можно частично синтезировать в машине, но дублировать с помощью ИИ программ его не удастся никогда.

В качестве отправной точки полезно рассмотреть дихотомию, предложенную философом из Беркли Джоном Сирлом (John Searl, 1980). Он описал две позиции в ИИ: "жесткую" и "мягкую"; согласно мягкой позиции, ИИ может использоваться как инструмент в исследованиях человеческого познания; а жесткая предполагает, что соответствующим образом запрограммированный компьютер обладает разумом и способен к пониманию. У "мягкого" ИИ мало оппонентов; почти все признают важность компьютеров в исследовании человеческого познания, и к этому почти нечего добавить. "Жесткий" ИИ, опровергаемый Сирлом, вызвал бурю протеста. Мы продолжим рассмотрение этого спора в следующем разделе про "Китайскую комнату", но сначала рассмотрим одну оригинальную задачу, предложенную британским математиком Аланом Тьюрингом⁴ и касающуюся разума и машин.

«Имитирующая игра» или «тест Тьюринга». Тьюринг предложил задачу, в которой человек задает вопросы неизвестному-существу-использующему-язык. Задача человека — решить, можно ли отличить *это* от человека. В пользу Тьюринга говорит то, что использование "имитирующей игры", ставшей впоследствии широко известной как «тест Тьюринга», само по себе было весьма тонким обманом, который, давая специалистам по ИИ нечто конкретное для работы, уводил их внимание от философских вопросов разума, ставших главным раздражающим фактором в истории науки и философии. Не обращаясь непосредственно к философским вопросам, как это сделал Тьюринг, он спрашивал: "Является ли познание функцией материальных процессов, и если да, то могут ли такие функции происходить от неорганической машины?" или "Как решить проблему тела и разума?" — т.е. он выбирал гораздо более четкие рамки вопроса, основанные на операционализме. Поскольку в литературе сохраняется определенная путаница относительно реальной природы предложенного Тьюрингом теста, мы приводим его здесь довольно детально. Эту... задачу можно описать в терминах игры, известной нам как "имитирующая игра". В нее играют три человека: мужчина (А), женщина (В) и спрашивающий (С), который может быть любого пола. Цель игры для спрашивающего —

определить, кто из двух других — мужчина, а кто — женщина. Для него они помечены как X и Y, и в конце игры он должен сказать либо: "X это A, а Y это B", либо: "X это B, а Y это A". Спрашивающий может задавать A и B такие, например вопросы:

C: Не скажет ли мне X длину своих волос? Теперь предположим, что X — это на самом деле A; тогда A должен ответить. Целью игры для A является... заставить C сделать ошибочную идентификацию. Его ответ мог бы быть таким:

"Мои волосы коротко острижены, самые длинные пряди длиной примерно 9 дюймов".

Чтобы тембр голоса не указывал спрашивающему на пол, ответы пишутся или еще лучше печатаются. Самый лучший вариант — это принтер, подключенный в соседней комнате. Другим вариантом может быть повторение ответов посредником. Цель третьего игрока (B) - помочь спрашивающему. Возможно, наилучшая стратегия для нее — это давать правдивые ответы. Она может прибавлять к своим ответам что-то вроде "Это я — женщина, не слушай его!", но это ничего не даст, поскольку аналогичные замечания может давать и мужчина.

Теперь спросим: "Что произойдет, если роль A в этой игре будет исполнять машина?" Будет ли спрашивающий при таком варианте игры ошибаться так же часто, как и тогда, когда играют мужчина и женщина? Эти вопросы заменяют первоначальный, "Может ли машина мыслить?" (Turing, 1950; p.434).

Очевидно, ценность некоторых вопросов, задаваемых для X и Y, зависит от того, какая сейчас мода, — т.е. если длину волос и прическу взять за основу различения, то, например, в 70-х годах это привело бы к очень многим ошибкам. Тем не менее, для специалистов по языку и ИИ в задаче Тьюринга есть очень важный момент— для того, чтобы компьютер перехитрил нас и заставил думать, что он — это человек, он должен понимать и генерировать ответную реакцию, которая эффективно имитировала бы важную когнитивную функцию.

Можно предложить и другой тип вопроса о неразличимости функций. Предположим, что в больнице работают два хирурга. Один — выпускник знаменитой медицинской школы и считается одним из лучших хирургов в мире Другой окончил обычную медицинскую школу и считается плохим хирургом. Однажды потребовалась срочная операция, и первый хирург нездоров, так что операцию проводит второй врач, о чем не известно пациенту, находящемуся без сознания. Пациенту не говорят, какой врач проводил операцию, и он рад, что она была успешной. К тому же, другие врачи уверены, что операцию проводил первый хирург. На этом ограниченном примере мы можем заключить, что тест на неразличимость был пройден. Однако, если бы вы были этим пациентом и узнали, что операцию на самом деле проводил робот, какой вывод вы бы сделали о соотношении функциональных свойств робота и хирурга? Согласились бы вы, что они одинаковы? Почему? А Почему нет? Ответы на эти вопросы трудно найти, чего не

скажешь о людях, придерживающихся твердых взглядов на этот счет Один из них — это Сирл, который вывернул "тест Тьюринга" наизнанку.

"Китайская комната". Чтобы продемонстрировать, что жесткая позиция в ИИ не выдерживает критики, Сирл предложил следующую головоломку. Предположим, что кого-то заперли в комнате, где много китайских текстов. Этот кто-то ничего не понимает по-китайски и даже не способен отличить китайские иероглифы от каких-нибудь еще. Снаружи этой комнаты ему передают еще один набор китайских знаков вместе с набором правил для сопоставления первого и второго набора символов. Эти правила всего лишь позволяют этому человеку связывать один набор символов с другим и написаны на обычном английском. При помощи этих правил сопоставления человек в китайской комнате может давать осмысленные ответы на вопросы о содержании китайских текстов, несмотря на то, что он в сущности не знает этого языка Через какое-то время эта личность приобретает настолько хорошую сноровку, что может отвечать как на своем родном английском, так и на китайском языке, которого не знает, но отвечает, основываясь на правилах Результаты настолько хороши, что их "совершенно нельзя отличить от ответов урожденного китайца" (Searl, 1980, 1981) Личность, запертая в китайской комнате, — это простая конкретная иллюстрация компьютерной программы "данные на входе — данные на выходе" Вплоть до этого момента почти никто из ИИ-пехоты не взъерошил перья, но затем Сирл шагает еще на один аргумент вперед Способность выполнять такие функции, как перевод по сложным правилам, не означает, что тот, кто это делает, понимает значение "выходных данных" Человеческий разум обладает *произвольностью* (см Searl, 1983), которая, согласно этому автору, определяется как свойство мысленных состояний и событий, направляющее их на объекты и ситуации в окружающем мире К таковым относятся убеждения, страхи, желания и намерения Независимо от того, насколько "неотличимо" поддельное мышление от "настоящего" (человеческого), они не есть одно и то же, поскольку у мыслящего человека есть намерения и поскольку между этими двумя "мыслящими" есть физические различия одного сделали органическим способом, а второго — электронным

Опровержение «китайской комнаты». Компьютерные ученые немедленно выдвинули возражения против головоломки Сирла, — прежде всего с позиций семантики термины "произвольность", "понимание", "мышление" употребляются им без четких операциональных референтов, в качестве возражения был приведен такой пример:

Если бы человек в этой "китайской комнате" выполнял описанные функции, то он (или система) действительно достиг бы как минимум некоторого уровня понимания Кроме того, аргументы Сирла отклонялись на основании 'приведения к абсурду' если довести ситуацию до логического конца, то оказалось бы возможным создать работа, в каждой своей детали идентичного мыслящему человеку, и все же последний был бы способен к "пониманию" и "произвольности", а первый нет Наконец, некоторые специалисты по ИИ полагают, что "понимание" и "произвольность" вызываются конкретными материальными свойствами Пилишин (Pylyshyn, 1980) сатирически замечает, что

возможно произвольность — это такое вещество, которое выделяется человеческим мозгом, он предлагает свою собственную загадку:

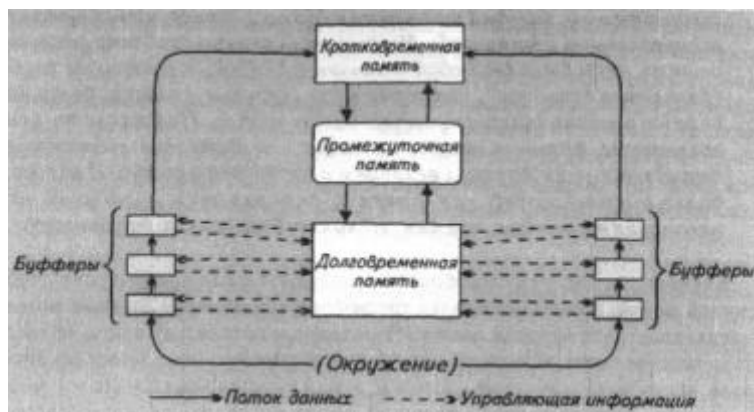
"Если бы все больше и больше клеток вашего мозга заменялись интегральными микросхемами, запрограммированными так, чтобы их характеристики входа-выхода были идентичны заменяемому элементу, вы по всей вероятности сохранили бы способность говорить точно так же, как и сейчас, за исключением того, что постепенно перестали бы что-либо под этим иметь в виду То, что мы, сторонние наблюдатели, все еще принимали бы за слова, для вас стало бы просто некоторым шумом, который заставляют вас издавать ваши микросхемы" (с. 442)

Этот спор далек от завершения, и некоторые, видимо, находят определенную ценность в его философской глубине. Однако, для меня этот спор не разрешен (на самом деле, он, вероятно, неразрешим!). Кроме того, оба лагеря ужесточили свои позиции и выдвигают скорее аргументы веры, чем разума. Для такой книги, как эта, его важность двояка: во-первых, он заставляет читателя глубоко задуматься о том "человеческом", что заключено в человеческом познании. Во-вторых, в связи с этим спором встает вопрос, до какого предела ИИ может имитировать человеческий интеллект. Страсти, разгоревшиеся вокруг "теста Тьюринга" и "китайской комнаты" у обеих сторон, отражают сильную заинтересованность современных философов и специалистов по ИИ в отношении электронного джина, выпущенного из бутылки.

КАКОГО РОДА КОМПЬЮТЕРОМ ЯВЛЯЕТСЯ ЧЕЛОВЕК?

Более функциональный подход, помещающий ИИ на когнитивную почву, предложил Ирл Хант (Earl Hunt, 1971); он поставил вопрос: "Какого рода компьютером является человек?" Исходя из того, что между обработкой информации человеком и компьютерной системой действительно существует аналогия, Хант предпринял грандиозную задачу описать компьютерную систему, которая "мыслит как человек". Фундаментальное положение его системы, названной им моделью распределенной памяти, гласит, что мозг имеет несколько зон памяти, назначение которых — регулировать поток информации из окружения так, чтобы информация могла кодироваться в терминах прошлого опыта. В основной структуре этой модели, изображенной (в модифицированном виде) на Рис. 15.1, информация из окружения проходит через ряд

буферных памяти и поступает в КП, или сознательную память.



Как показано на рисунке, одновременно может обрабатываться более, чем один сенсорный канал. Буферная обработка служит для преобразования

сообщений в постепенно более значимые единицы. Это можно проиллюстрировать на примере чтения, где сначала различаются признаки (в ранних буферах), затем буквы и, наконец, слова; единственный сознательный процесс (у опытного читателя) — это собственно чтение слов. Центральная роль в этом процессе принадлежит ДП, в которой информация хранится постоянно. Кодирование сырой информации и ее перекодирование на более высоких уровнях происходит под управлением КП. Вербальная информация, достигающая КП, интерпретируется в значимый семантический код, хранимый в промежуточной памяти (ПП). Этот семантический код сохраняет непосредственный опыт, так что мы можем изучать его дальше, извлекая из него информацию и интерпретируя ее. Общая архитектура этой системы, как отмечает сам Хант, носит поразительное сходство с некоторыми из моделей памяти, с которыми мы встречались в Главе 5 — особенно с моделью Аткинсона и Шифрина.

Таким образом, модель Ханта может служить руководством по конструированию моделей переработки информации у человека, причем для имитации конкретных когнитивных функций проектируются конкретные компьютерные программы. Возможно, что основным результатом процесса моделирования является возможность испытать наши знания о когнитивной природе человека. Компьютерные программы, которым не удастся точно смоделировать человеческое познание (и часто это так и есть) могут иметь большее значение для определения единственно человеческих свойств человеческого познания, чем некоторые преуспевающие программы. До сих пор мы в общем наметили сферу ИИ и познакомились с постулатом человеческого познания, на основе которого можно[^] осуществлять компьютерное моделирование. В следующем разделе мы разберем некоторые конкретные возможности компьютера. Развитие этих конкретных функций в модели на основе информационного подхода примерно соответствует потоку информации от восприятия к распознаванию паттернов и высшим формам познания.