

Автор: Петров И. Б.

Дата: 17.08.2019

Св-во о публикации: №219081701432.

E-mail: petrovibmath@yandex.ru

Принцип песочных часов или истинное лицо разумного интеллекта.

Искусственный интеллект (далее по тексту – ИИ) давно стал вожделенной целью многих ученых, специалистов в области информационных технологий, футурологов и футуристов, а также всех тех людей кто верит в научно-технический прогресс будущего человечества. Однако не смотря на множество попыток создать ИИ, множество теорий его описывающих, а также не меньшее количество подходов к этой, уже кажется извечной проблеме, – создать его пока никому так и не удалось... Все попытки поспорить с этим утверждением упираются в упрямые догмы определений самого понятия ИИ. Еще в 1980 году известный американский философ Джон Серл ввел понятие «сильного ИИ», как ИИ аналогичного человеческому разуму, то есть такого механизма, который должен обладать самосознанием (некой личностью способной осознавать свои мысли) и умением решать и обосновывать (решения) различных задач.

На сегодняшний день существует множество ИИ (в том числе применяемые во многих областях экспертные системы), которые в полной мере обладают способностью решать определенный круг задач (включая симуляцию человеческого поведения и речи), однако не способны ни коим образом осознавать свои собственные решения. По сути все эти ИИ – сложные аппаратно-программные комплексы (АПК), на самом деле своей сутью представляют всего лишь автоматы, самый простейший из которых человечество изобрело на заре эры информационных технологий и дало название – калькулятор. Не следует думать, что такое сравнение – это попытка как-либо принизить трубу многих поколений замечательных ученых, конструкторов, математиков и программистов. Нет, все эти АПК – необходимая ветвь развития науки, определяемая практической потребностью современного общества, с которой все эти решения прекрасно справляются, однако, СИИ они не являются по своей сути и определению.

Но в чем же проблема создания СИИ? И ответ на этот вопрос кроется в непонимании собственного (человеческого) мышления. В особенности, той его части, которая называется – сознанием и самосознанием. Первый термин скрывает в себе способность субъекта представлять окружающий мир и позиционировать себя в нем (кто я по сравнению с окружающими объектами? И где я в этом мире нахожусь?). А вот со вторым понятием все намного сложнее: самосознание – это способность абстрагировать (рассматривать) самого себя как отдельного субъекта окружающего

мира (то есть понимать, что я – есть разумное) и самое главное – понимать свои собственные мысли (то есть осознавать, что я способен мыслить и все решения принимаются именно мной). Если не углубятся в научные и философские термины, то самосознание включает в себя все то, что любой из нас подразумевает под собственным «Я». И вот на этом моменте, при создании СИИ, возникает две огромные проблемы: как заставить машину (автомат, ИИ) осознавать себя в окружающем мире и сознать себя некой личностью?

Первую проблему, вероятнее всего, решает теория модельной природы мышления, которая заключается в том, что наш мозг просто моделирует окружающее (нас) пространство исходя из накопленных знаний о нем. Представьте ситуацию: вы просто идете по улице... Вы осознаете себя самим собой и позиционируете себя в пространстве этой улицы (понимаете где вы находитесь), а также способны абстрагировать (отделять) себя и свой разум от других прохожих, понимая – что вы отдельный человек. В этот момент, следуя модельной теории в вашем мозгу, как в некой компьютерной виртуальной реальности (например, как в компьютерной игре) моделируется окружающее вас пространство (строятся виртуальные модели того, что видят ваши глаза), и при этом каждому моделируемому объекту задаются параметры (физические, смысловые, эмоциональные и т. п.) исходя из данных содержащихся в вашей личной базе знаний (накопленной в течении практического опыта и поступившей информации), которая продолжает пополняться, как только вам станет доступно для понимания (обработка новой информации) новый, ранее не известный, параметр того или иного объекта. Такое описание кажется сложным, но попробуйте представить в ключе модельной теории вашу повседневную жизнь. Например, что вы знаете о обычно кирпиче? Прежде всего то, что он тяжелый и твердый, из него строят здания и он может представлять опасность – упав вам на голову. Но заметьте, еще до такого как вы мысленно перечислили некоторые свойства (параметры) кирпича, вы его представили в собственной голове. Это и есть модель. Если мы говорим о знакомом нам предмете (или даже человеку), мы всегда представляем его внешне или по содержанию (смыслу), то есть строим модель. Такое небольшое замечание отчасти доказывает верность теории модельной природы мышления, но не объясняет как мозг формирует второе – самосознание. Ведь представив и описав свойства кирпича, каждый из нас можем с уверенностью сказать, что сейчас думает о кирпиче (то есть понимает, что в его (личном) мозгу сформировалась конкретная (персональная) мысль). И каждый из нас легко можем смоделировать применение кирпича как для строительства дома, так и для раскалывания орехов, полностью понимая и осознавая, что именно это и именно он делает, взяв в руку этот кирпич... И теории модельного мышления явно не достаточно, чтобы описать принцип самосознания. Ведь не возможно представить способность понимать свои мысли в виде некой модели. Вообще человеку довольно сложно описать этот принцип и главное осознать его, о чем

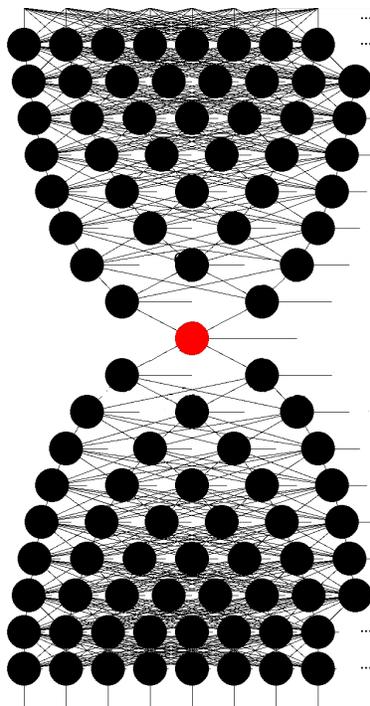
свидетельствуют множество значений данного термина, множество подходов и представлений его в различных психологических и философских школах. Но давайте не станем углубляться в рассуждения о самопознании, а остановимся на том факте, что наша задача в рамках вопроса поднятого в данной статье всего лишь – научить этой способности машину (автомат). При этом, мы должны достаточно хорошо понимать, что мышление машины в общем и сознание, самосознание в частности, могут весьма резко отличаться от аналогичных понятий человеческого мозга. И это вполне естественно, если только мы не хотим создать копию такого мозга. Для СИИ будет достаточно обладать лишь аналогичными механизмами по своему принципу, но совершенно не обязательно, чтобы эти механизмы имели схожую техническую (логическую, алгоритмическую) реализацию.

И одной из таких логических реализаций может быть «принцип песочных часов», который в данном контексте сводится к подобной (логической) схеме: представьте себе самые обычные песочные часы – в верхней камере находится песок и проходя по одной песчинки через заужение в сосуде, он попадает в аналогичную камеру внизу. Представим, что все песчинки в верхней камере символизируют собой входные сигналы, то есть поступающую информацию из окружающего мира. Так, любая информация с которой соприкасается СИИ словно песчинки группируются и собираются всей своей массой в зауженном месте сосуда, которое означает собой некую «точку сборки», по сути являющимся неким аналогом значения человеческого «я» (в том смысле, что это внутренняя позиционная точка, от которой возможно абстрагироваться от окружающего мира и внутренней структуры разума (всего остального объема песчинок), необходимая для осознания мыслительного процесса, другими словами: точка, относительно которой мы можем наблюдать свои же мысли). Проходя через эту точку, каждый входной сигнал поступающей информации (песчинка) обрабатывается, относительно точки своего «я» и поступает в виде выходных сигналов в нижнюю колбу. При этом важно понимать, что количество песчинок в верхней колбе, в итоге всегда будет равно (после обработки всей поступившей информации) количеству оных в нижней. На этом важном для понимания описываемого принципа моменте хочется остановиться немного подробнее. Дело в том, что входящая информация о каком-то конкретном предмете или явлении окружающего нас мира имеет определенное количество входных сигналов. Вероятнее всего, оно равно количеству рецепторов органов чувств, которые взаимодействует с данным предметом в данный момент времени или в конкретной ситуации.

Для простоты представления возьмем упоминаемый ранее кирпич. Допустим для его описания в рамках СИИ (представления и осознания его формы и свойств – параметров базы знаний) необходима сотня датчиков-рецепторов. Значит на входе мы будем иметь сто сигналов. Однако и на выходе, обработав при помощи СИИ

информацию о кирпиче, мы также должны выдать сто сигналов. Дело все в том, что это необходимо по двум причинам: первая из них – это моделирование объекта (кирпича) в виртуальном мозге СИИ, которое представляет собой обыкновенное проецирование на виртуальное пространство. Для этого нам понадобится равно такое же количество критериев (параметров) объекта, как и для его восприятия органами чувств – ведь мы переносим все доступные нам свойства объекта, как бы копируя и воссоздавая его в нашем мозгу. Если Природа и не пошла по принципу наименьшей сложности в данном механизме, то нам для реализации СИИ будет проще именно такой принцип схемы: сколько входов – столько и выходов. Второй же причиной данного принципа является то, что будь-то человек, будь-то СИИ, ответная реакция на предмет (действие над ним) будет выражаться все через тоже количество рецепторов. Допустим мы идем по улице и видим на земле кирпич. Как только мы его увидели рецепторами глаз: в нашем мозгу сформировался виртуальный объект – тот самый кирпич, при том состоящий из такого же количества точек (условно назовем так), которыми мы этот кирпич увидели. И теперь, если наш разум примет решение обойти кирпич так, чтобы не споткнуться об него, то сформировав траекторию, реализуемую двигательными органами, мы будем наблюдать во время нашего движения за данным кирпичом все тем же количеством зрительных рецепторов, формируя ответную реакцию по крайней мере на габариты данного кирпича, стараясь не зацепить его своими ногами. Такая схема подходит для реализации СИИ с практической точки зрения.

Теперь попытаемся разобраться с собственным «Я» нашего будущего СИИ. Можно долго рассуждать и безрезультатно спорить о том, на сколько сложно устроило «Я» в человеческом разуме Природа, но одной из логических реализаций на машинном уровне является сведение всей сети обработки информации к одному ее элементу. На примере нейроподобных сетей – это означает сведение к одному нейрону. Давайте для наглядности рассмотрим такую логическую (не практическую) схему нейроподобной сети:



Сверху находятся синапсы (канал связи между двумя нейронами) подводящие к данной сети сигналы от внешних рецепторов. Рассматриваемая сеть имеет несколько уровней нейронов, каждый из которых содержит меньшее их количество по сравнению с предыдущим, таким образом, что аксоны (выход нейрона) всех элементов сети сводятся к одному единственному нейрону (обозначенному на схеме красным цветом), который и характеризует структуру ответственную за собственное «Я» нашего СИИ. Представим, что рецепторы СИИ наблюдают за все тем же кирпичом, с целью его преодолеть. Тогда, вся масса информации о лежащим на дороге кирпиче сведется к одному единственному нейрону, вычислитель передаточной функции (вычислитель значения выходного сигнала) которого будет определять принятие решения о моделировании данного кирпича с целью дальнейшего формирования действия над ним (в данном случае действие по преодолению его). Вес сигнала подаваемый на аксон нашего нейрона фактически будет определять значение весов выходных сигналов, которые сформируются при развертке сигнала этого нейрона по второй половине (нижней на схеме) сети. Если допустим мы описываем кирпич сотней входных сигналов, то и на выходе мы получаем ту же сотню сигналов (но с иными значениями).

Представим мозг нашего СИИ как множество подобных «песочных часов», выходы одних из которых являются входами других, таким образом, что все «часы» соединены между собой и последовательно, и параллельно. Однако центральные («красные») нейроны соединены синапсами напрямую только между собой, образуя таким образом как бы свою нейронную сеть в общей сети. Именно эта сеть и является механизмом самосознания СИИ, так как она может реагировать на потоки (мысли) текущей информации (проще говоря – ее входными сигналами являются наличие потоков линейной информации, проходящими через каждые «песочные часы»). Таким образом каждая сеть по типу «песочных часов» - обрабатывает конкретную информацию (отвечает только за определенную малую задачу), в то время как сеть сформированная из «красных» нейронов обрабатывает «информацию об информации», косвенно управляя каждым «песочными часами» и могущая формировать независимый от них поток информации (характеризующий личность СИИ – то есть его «Я»). Такая сеть по своему определению одновременно будет независима от внешней информационной среды (будет самостоятельно формировать потоки информации), но при этом иметь ответную реакцию на любые входящие сигналы из окружающего мира. Что если не это можно назвать – личностью?

Остается открытым лишь следующий вопрос: а нужна ли личность СИИ? Но ответ на него весьма прост – без наличия у СИИ своего собственного «Я» не возможно создать полностью самостоятельный механизм. Представим ситуацию, пусть и

футуристического толка, но зато отлично демонстрирующую необходимость такой самостоятельности у разумных машин: допустим в будущем вам необходимо создать искусственного спутника (например, друга или компаньона) для человека в его повседневной жизни. Ваш робот обладает всеми необходимыми аппаратными механизмами (телом, датчиками и т. п.) и может выполнять свои функции (это доступно уже сегодня). Казалось бы достаточно прописать алгоритмы обучения (пополнение базы знаний) и ответных реакций на события окружающей среды... Но проблема в том, что: во-первых, всего предусмотреть невозможно, а во-вторых – если у вашего робота нет СИИ, то вам придется программировать огромное множество возможных вариантов развития событий даже на один из критериев окружающей среды (а сколько таких критериев всего?!). Другое дело, если речь идет о полноценном разуме личности, которая будет обладать ценнейшей способностью самостоятельно принимать решения, даже в неучтенных программистами ситуациях. Более того – отпадет необходимость в программировании большей части алгоритмов ответных реакций на те или иные ситуации.

Конечно, можно долго говорить об опасности такой машины (например, для человека – как вида разумного), но если задуматься найдется не один десяток сфер деятельности для современных ИИ (медицина (независимые медицинские автоматизированные комплексы), психология (персональные собеседники и партнеры), универсальные помощники-консультанты и т.п.), где такая самостоятельность крайне необходима. А безопасность может гарантироваться искусственным формированием «характера» машины (ведь прежде чем запускать такой отдельный СИИ, его «личность» может заранее программироваться под необходимые нужды).

Таким образом, в конце этой относительно небольшой и несложной по содержанию информации статьи, хочется отметить, что приведенная гипотетическая модель СИИ, возможно, не является чем-то в особенности новыми и уникальным, тем более не конкретизирует схему по его создания, но тем ни менее может служить отправной точкой для нового взгляда на проблематику создания СИИ, ИИ и ЭС нового поколения. А также подчеркивает возможность использования нейроподобных сетей для создания таких разумных автоматов.