

Когда начинать бояться ИИ?

Часть 2. ИИ или Искинт?

Оглавление

КАК НАЧИНАЛСЯ «МАШИННЫЙ РАЗУМ»?	2
ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СССР	4
КАРТИНА МИРА.	11
«РАБ» ИЛИ «ДРУГ»?	14
ЛИТЕРАТУРА:	16

А вообще, что такое этот Искусственный Интеллект?

Или может быть это Машинный Разум?

Почему «машинный разум»? Мы же об искусственном интеллекте...?

Я не знаю, когда началось это разночтение одних и тех же терминов у нас и на Западе. Видимо, давно. Одно и то же слово у нас и в западном понимании – разные понятия. Даже самые важные для человека.

Например, «разум» и «интеллект».

Для нас разум, это важнейшая характеристика человека. Ну конечно, лат. Homo sapiens, это же «человек разумный», и у нас разум – «способность мыслить всеобще, способность анализа, абстрагирования и обобщения.» <https://ru.wikipedia.org/?curid=122626&oldid=128547235>

В западном понимании всё иначе.

«Латинское слово ratio также обозначает разряд, рассудок, рассуждение, расчёт, что служит иллюстрацией существовавшего некогда объединения понятий разума и рассудка. Но рассудок в современных представлениях не создаёт нового знания, а лишь систематизирует уже существующее.» <https://ru.wikipedia.org/?curid=122626&oldid=128547235>

А вот с интеллектом всё наоборот, на западе он эквивалентен нашему «разуму»... Здесь интеллект - *«(от лат. intellectus «восприятие»; «разумение», «понимание»; «понятие», «рассудок» или ум — качество психики, состоящее из способности осознавать новые ситуации, способности к обучению и запоминанию на основе опыта, пониманию и применению абстрактных концепций, и использованию своих знаний для управления окружающей человека средой.»* <https://ru.wikipedia.org/?curid=3740&oldid=129152652>

И потому, всё, что мы пишем о «разуме» человека, на западе написано про «интеллект», а то, что мы понимаем, как «интеллект», там определяет «разум».

Сегодня я вполне допускаю, что это делалось когда-то совсем не случайно, а очень даже намеренно.

Но...

В общем случае на Западе и в СССР на первом этапе понимания этой проблемы все говорили и думали об одном и том же. О «машинном разуме».

Ученые обсуждали именно «разум».

Как начинался «машинный разум»?

Мечта о «машинном разуме» стала конкретизироваться в рамках мирового движения науки к «искусственному человеку», начатой еще в начале прошлого века. Тогда казалось, человек может всё. Может скрестить человека, какую-нибудь гориллу и получить «идеального солдата», или взять кучу реле, электродвигателей и радиоламп и получить «железного человека» с «электронными мозгами», который будет солдатом и рабочим не хуже человека. Термин «робот», использованный К. Чапеком в 1920 году для таких «железных» «рабов» вполне подходил и по звучанию, и по содержанию.

В это же время стали появляться и настоящие логические машины, которые почти автоматически получали ответы по логическим правилам, заложенным в машину. А чуть позже появились и первые, сначала электромеханические, а потом электронные вычислительные машины, которые считали в тысячи раз быстрее человека.

«Спусковым крючком» начала этого лавинного процесса стала новая наука об управлении - кибернетика. Даже только своим появлением, она вдруг показала реальный путь к цели – машинному разуму, о котором вся наука мечтала уже полвека.

В середине прошлого века «центр мира» находился на Западе. Вычислительная машина, программирование и кибернетические принципы управления создали базис для запуска там процесса обсуждения путей реализации «машинного мозга», содержащего «машинный разум», аналогичный человеческому.

Потом, очень скоро, этому воспротивились западные математики.

Они разделили для себя понятия «интеллект», как способности к комбинаторике, и «разум», как способности к абстракции и обобщению.

И теперь есть такое понимание:

«...В психологии и нейробиологии спорная теория множественного интеллекта применяет термины интеллект (эмоции) и интеллект (разум) для описания того, как люди понимают мир и реальность.

... интеллект относится к логическим и рациональным функциям человеческого разума и обычно ограничивается фактами и знаниями.

*...Интеллект и интеллект противопоставляются по этимологии; термин *intelligence*, образованный от латинского настоящего активного причастия *intelligere*, означает “собирать между”, тогда как термин *intelligence*, образованный от причастия прошедшего времени *intelligere*, обозначает “то, что было собрано”. Следовательно, интеллект относится к созданию новых категорий понимания, основанных на сходствах и различиях, в то время как интеллект относится к пониманию существующих категорий.»*
<https://en.wikipedia.org/wiki/Intellect>

Ну как-то так...

Таким простым и эффективным ходом западная наука отделила себя от философских проблем постижения «машинного интеллекта» и занялась его научным рассмотрением.

Более или менее научное осознание проблемы ИИ происходило на конференциях Мейси, начиная с 1941 и заканчивая 1960г. [1], и закончилось формулированием научного определения ИИ, которое ничего общего не имело с тем, что рассматривали первые философы ИИ. Искусственный интеллект, как *свойство искусственных интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека; наука и*

технология создания интеллектуальных машин, особенно интеллектуальных компьютерных программ ...в виде проблемы, которую предстоит решить науке впервые была рассмотрена в далеком 1956г на семинаре в Дартмутском университете:

«Согласно Маккарти, исследователи вольны использовать методы, которые не наблюдаются у людей, если это необходимо для решения конкретных проблем. Поясняя своё определение, Джон Маккарти указывает: «Проблема состоит в том, что пока мы не можем в целом определить, какие вычислительные процедуры мы хотим называть интеллектуальными. Мы понимаем некоторые механизмы интеллекта и не понимаем остальные. Поэтому под интеллектом в пределах этой науки понимается только вычислительная составляющая способности достигать целей в мире».» <https://ru.wikipedia.org/?curid=2665&oldid=128557610>

Поразительно, но основная масса ученых в понимании основы искусственного интеллекта так и идет в направлении, указанном Джоном Маккарти.

А тем временем...

Зарождалась западная философия ИИ, вместе с развитием кибернетики и последующим переходом её основного направления развития с логической машины на вычислительную [2]. Основы были заложены не философами, а писателями-фантастами А.Азимовым (законы роботехники) 1942г. и С.Лемом (Сумма технологии) в 1963г. Первый своими «законами» отделил интеллект робота от человеческого интеллекта, второй противопоставил. Это определило характер дальнейшего разговора о понимании искусственного интеллекта. Его сравнивали с человеческим и противопоставляли.

Техническое и философское направления разошлись во времени и пространстве. Научное, математическое и техническое направление развития ИИ в западных науках окончательно перестало обращать внимание на философские обсуждения этих проблем. Более того, западная наука начала писать новую философию развития ИИ.

В послесловии к книге Х. Дрейфуса [3] Б.В.Бирюков пишет:

«Какова же на самом деле та философская традиция, на основе которой наука пришла к кибернетике, к моделированию познавательных процессов, к направлению "искусственного интеллекта"? Прежде всего это логическая традиция — традиция, исторически восходящая к Лейбницу.

...В числе предшественников современных модельно-кибернетических исследований, помимо Лейбница и Буля (о которых говорится в книге), следует назвать также и выдающихся математических логиков конца прошлого века Э. Шрёдера и П. С. Порецкого (Шрёдер вполне определенно сформулировал идею "искусственного интеллекта"), и Г. Фреге, выдвинувшего программу логической формализации математики, и Д. Гильберта, положившего начало современным метаматематическим исследованиям, и К. Геделя, А. Тьюринга, Э. Поста, С. Клини, А. Тарского, А. Маркова, П. Новикова и др., заложивших основы теории алгоритмов и в серии знаменитых теорем раскрывших внутреннюю ограниченность формализации.» [3]

Точно и емко указана та первая ветвь развития ИИ, которая первой и пришла к своему тупику...

«Символический искусственный интеллект — собирательное название для всех методов исследования искусственного интеллекта, основанных на высокоуровневом «символическом» (человекочитаемом) представлении задач,

логики и поиска. Символический ИИ лёг в основу доминирующей парадигмы исследований ИИ с середины 1950-х до конца 1980-х. В 1985 году Джон Хогланд (англ. John Haugeland) дал символическому ИИ название GOF AI (англ. Good Old-Fashioned Artificial Intelligence, «старый добрый искусственный интеллект») в своей книге *Artificial Intelligence: The Very Idea*, посвящённой философскому отражению последствий исследований искусственного интеллекта. В робототехнике применяется аналогичный термин GOF AIR («старый добрый искусственный интеллект в робототехнике»).

<https://ru.wikipedia.org/?curid=6178634&oldid=127563354>

Good Old-Fashioned Artificial Intelligence, «добрый старый искусственный интеллект», так это переводится, «ушел в тень» в начале 80-х...

В своей работе [2] я когда-то написал: *Формально, в 70-80 гг. прошлого века было окончательно установлено, что уровень развития вычислительной техники не позволяет говорить даже о возможности приближения к Искусственному Разуму. «Разумность» автоматических систем была снята с рассмотрения.*

Это было написано после прочтения многих публикаций по кризису ИИ в то время. Ну ладно, а что сегодня?

Наиболее успешная форма символического ИИ — это экспертные системы, использующие сеть продукционных правил. Продукционные правила объединяют символы в отношения, похожие на оператор «если-то». Экспертная система, обрабатывая эти правила, делает логические выводы и определяет, какая дополнительная информация ей необходима, то есть какие следует задать вопросы, используя человекочитаемые символы.

...Однако на практике системы, основанные на данных принципах, хотя и работали, но плохо справлялись со сложными адаптивными задачами. Поэтому в 1980-90-е годы такие концепции подверглись серьёзной критике, и интерес многих исследователей сместился в сторону других методов (эволюционные алгоритмы, искусственные нейросети и т. д.)»
<https://ru.wikipedia.org/?curid=6178634&oldid=127563354>

Вот так сегодня наука уже в справочниках подтверждает мои слова.

Тогда же, с середины прошлого века вся мировая наука, а вместе с ней и советская, шли по этому пути строить искусственный интеллект. На этом направлении развития ИИ сосредоточилась западная и советская философская мысль. В рамках этого направления кипели споры, происходили и настоящие сражения противоборствующих сторон, рождались новые гипотезы...

Это произошло примерно к 70-м годам прошлого века.

Философия искусственного интеллекта в СССР

С середины 60-х инициативу в философском и техническом понимании ИИ перехватил СССР. Теперь здесь стали собираться международные конференции по ИИ с участием ученых разных специальностей и выпускаться ежегодники с публикацией статей по этой тематике [4]. Основоположителем ИИ в СССР, а потом и в России, был Д.А.Поспелов.

Тогда, в середине прошлого века вперед сразу вышло направление символического программного интеллекта, основанного на логических правилах, командах и их исполнении машиной.

Успех был потрясающим.

Следует напомнить, что пионерская работа К. Хьюитта в области теории агентов «Viewing Control Structures as Patterns of Message Passing» вышла в свет в 1977 году тогда как первые публикации Д.А.Поспелова по психонике и модели агента- гиромата, появились на десять лет раньше.

На заседаниях семинара, проходивших в 1967-1968 г. сформировались задачи, подходы и методы, составившие основу ситуационного управления. В нём нашли прямое отражение идеи гироматов.»[7]

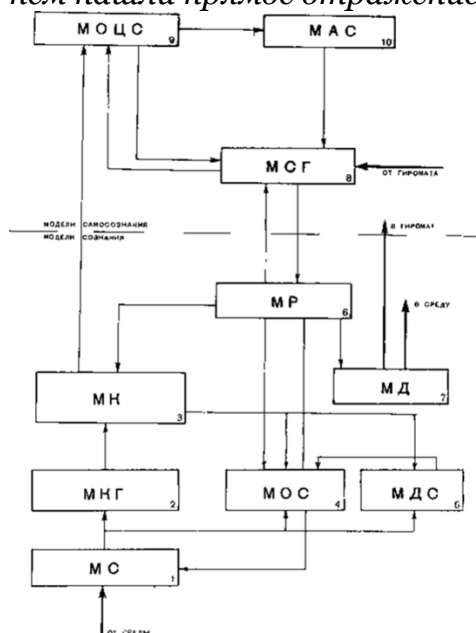


Рис.2. Система «думающей» машины с применением гиромата.

отображения этого мира в памяти машины связана с необходимостью отображения его структуры, его динамики и законов взаимодействия мира с машиной.

...В настоящей работе мы покажем преимущества гироматов по сравнению с обычными вычислительными машинами, в которых полностью игнорируется анализ исходной задачи из-за отсутствия средств для такого анализа. Это преимущество будет продемонстрировано на примере таких категорий, как «сознание» и «самосознание», интерпретированных для обычных вычислительных машин и гироматов. ...Будем в дальнейшем условно называть «сознанием» гиромата его свойство отображать внешнюю среду в своей памяти и анализировать закономерности этой среды и результаты своих воздействий на среду. Под «самосознанием» гиромата будем условно понимать его свойство отображать себя в модели среды и анализировать закономерности воздействия среды на свою структуру и функционирование.

...Дадим краткие пояснения по функционированию системы моделей гиромата. Источником информации для гиромата служит вход от внешней среды, по которому в модель ситуаций (1) поступают сигналы от внешних органов чувств гиромата, и вход от некоторой части устройств самого гиромата (от датчиков на этих устройствах, играющих роль внутренних органов чувств гиромата).

Сигналы от внутренних органов чувств поступают на модель структуры гиромата (8), которая сама представляет собой совокупность моделей со структурой, подобной структуре моделей сознания на рисунке. Сигналы, поступающие по любому из этих входов, являются

Всё так, но...

Исходно идея гиромата должна была привести к «разуму» машины, который бы соответствовал его человеческому пониманию. Гиромат должен был стать «мозгом» машины. Он должен был обладать сознанием и учитывать мораль и этику в принятии решений...

Гиромат должен был как-то соотноситься в функциональной схеме своей работы с хоть примерной, но «схемой» человеческого мозга. Это и отражает схема «думающей машины» на рис.2.

«Единственной возможностью расширения способностей машины к решению творческих задач, адаптации в неисследованной среде, самоорганизации и т. д. является создание внутри нее семиотической системы ее внешнего мира (например, мира задач). ...Сложность

структурированными сигналами, а не кодовыми идентификаторами, как это принято в обычных вычислительных машинах.» [5]

Как показало время, идея оказалась очень сложно реализуемой. Направление было взято верное, но в то время еще не было тех знаний, которых, кстати, нет и сейчас.

И тем не менее Д.А.Поспелов упорно идет по намеченному пути...

Чуть позже он разрабатывает основы психоники, раздела бионики, использующего результаты психофизиологических исследований при построении систем управления.

О психонике только цитаты:

«Так с «лёгкой руки» Д. А. Поспелова родилась психоника -- научная область, основной задачей которой стало изучение и использование в интеллектуальных системах результатов, добытых психологами при изучении психики человека и способов организации человеческой деятельности. На первом заседании семинара был прочитан доклад Д. А. Поспелова «О задачах психоники» (см. Поспелов Д. А. О задачах психоники. Проблемы бионики. М.:Наука, 1967, с.294-297.), в котором было отмечено появление круга проблем, которые с одной стороны являются предметом исследований психологов и лингвистов, а с другой стороны, представляют большой интерес для инженеров, специалистов в области технических наук, математиков, программистов, которые занимаются проектированием устройств, предназначенных для целенаправленной, «логической» деятельности. К таким проблемам относятся вопросы целенаправленного поведения, мотивации поступков, выработки абстрактных понятий, проведения индуктивных выводов, и т.д. Автором были указаны три направления в психонике:

- 1) создание в машине внутренней модели внешней среды;*
- 2) принятие целесообразных решений;*
- 3) моделирование личности и коллективов.*

Было отмечено, что надделение технических систем аналогами личностных характеристик позволило бы строить управляющие устройства, обладающие некоторой «субъективностью» подхода к решению задачи. Например, при моделировании игровой ситуации следует учитывать такие поведенческие аспекты как обман и блеф.» [8]

Семинары по психонике, гироматам и др. проводились Д.А.Поспеловым в МЭИ с 1964 по 1970г. За эти годы было проведено 92 семинара [9]:

«С самого начала в его заседаниях принимали активное участие не только математики, программисты и инженеры, но и психологи и лингвисты. Такой состав участников диктовался междисциплинарным характером исследований, связанных с созданием устройств психонического типа. Характеризуя особенности таких устройств в своем докладе "О задачах психоники", прочитанном на первом заседании семинара, Д.А.Поспелов отметил, что "основным отличием машин психонического типа от обычных вычислительных машин является наличие в них модели внешнего мира, которая отражает основные соотношения этого мира с помощью системы базовых понятий и отношений различного типа между этими понятиями". Таким образом, в машинах психонического типа явно предполагалось то, что позже стали называть базами знаний.» Д.А.Поспелов Семинар по психонике стр.31. [9]

И наконец, результат:

«Сам Д. А. Поспелов отмечал, что семинар по психонике в МЭИ стал одним из ключевых мероприятий, связанных с зарождением ИИ в нашей стране. Хотя в отличие от бионики, сам термин «психоника» так и не прижился в научной среде, сами исследования в этой области привели к формированию междисциплинарного научного сообщества, необходимого для развёртывания широкого фронта работ, относящихся к искусственному интеллекту. К этому можно добавить, что, начиная с середины 1960-х годов, сам Дмитрий Александрович выполнил немало экспериментальных исследований совместно с психологами. В первую очередь, искались альтернативы чисто бихевиористским подходам к решению задач, опиравшимся на лабиринтную модель. Именно результаты таких совместных исследований способствовали формированию широкого менталистского воззрения Д.А.Поспелова на предмет искусственного интеллекта, отнюдь не сводящегося к классической схеме инженерии знаний. В соответствии с идеями ментализма, адекватная характеристика сложного, целенаправленного человеческого поведения и самой жизни невозможна без привлечения сознания или психики в целом как «средств объяснения».» [4]

Финал и здесь оказался ожидаемым. Д.А.Поспелов в своих идеях обогнал время почти на век и результат тогда не был достигнут... не потому, что он шел не туда, а потому, что наука о мозге за ним не успевала...

Сегодня понятие «психоника» используется другими. Например, геотеизмом [10]. Но опыт понимания подходов к строительству Искинта никуда не делся.

«К числу новых научных направлений в области ИИ, авторство которых принадлежит Д. А. Поспелову, также относятся предложенные им **псевдофизические логики** (моделирующие рассуждения "здравого смысла" о времени, пространстве, действиях, причинно-следственных цепочках и т .п.), **формальные модели индивидуального и коллективного поведения, психоника** (наука о психологии искусственных систем), **когнитология** (междисциплинарное исследование процессов познания), **теория гиромата** (предвосхитившая многие идеи и положения современной теории агентов) и **модели децентрализованных** (а по сути, многоагентных) **систем.**» [8]

Вот здесь можно остановиться и задуматься...

Эйфория быстрого понимания и построения Искинта, аналогичного человеку по разуму, прошла. На том уровне техники и технологии проблема оказалась неразрешимой.

В книге математика В.Лефевра «Конфликтующие структуры» вышедшей в СССР в 1967г [11], рассматриваются уже несколько другие вопросы:

«Лефевр в своих работах начал проводить мысль, что естественнонаучный подход не являлся адекватным для описания человека. Традиционное противопоставление исследователя и объекта исследования теряло смысл при исследовании объектов, сравнимых с исследователем по совершенству, ибо они были способны проникать в замыслы исследователя и влиять на теорию, им создаваемую. Для изучения человека была необходима формальная модель, подобная физическим моделям. Человека следовало представлять неотъемлемой частью физического мира и с присущими ему феноменами сознания: образами себя и другого вместе с отображениями ими образов друг друга (Конфликтующие структуры, 1967).
<https://ru.wikipedia.org/?curid=26134&oldid=126952714>

Он также поставил вопрос о создании формальной модели человека, для сравнения параметров машины с этой моделью. В том числе и параметры интеллекта и разума. Правда потом сам В. Лефевр, переехав в США, ушел от этого понимания.

И все же...

О «разуме» машины думал не только Д.А.Поспелов.

В 1978г выходит книга А. Г. Чачко «Искусственный разум» Серия «Эврика» [12]:

«Сегодня в сложном термине «искусственный интеллект» куда ни кинь, всюду сомнения. Здесь, быть может, полезно сокращение: под его прессом отпадут ненужные смысловые оттенки, останется лишь существо дела. Давайте впредь пользоваться словом «Искинт», понимая под Искинтом уровень развития дела.

Книга, которая предлагается вашему вниманию, — повесть об Искинте. Рассказывая о сегодняшнем дне, нельзя упускать из виду дальнюю перспективу, которую очень четко определил академик В. Глушков: «Конечная цель... — создание искусственного интеллекта, не только не уступающего, но и намного превосходящего по своим возможностям естественный человеческий интеллект. Хотя эта цель пока еще достаточно далека, работа по ее достижению уже идет полным ходом. При этом ставятся вполне конкретные практические задачи повышения производительности и уровня интеллектуальности компьютеров для быстрого повышения производительности труда человека на многих участках интеллектуальной деятельности.»[12]

Может быть автор уже тогда немного запоздал со своей книгой, но так думали в СССР многие специалисты по Искинту. К началу 80-х в этой отрасли сложились научные школы разных направлений вычислительной техники, по-разному относящихся к пониманию искусственного интеллекта, его философии и целей. Техника уже явно не успевала за идеями...

А что в это время происходило на Западе?

Вопрос «Может ли машина мыслить?», поставленный А.Тьюрингом в 1950г стал главным вопросом философии ИИ. Одной из самых заметных в то время стала книга Х.Дрейфуса ЧЕГО НЕ МОГУТ ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ. КРИТИКА ИСКУССТВЕННОГО РАЗУМА [3], опубликованная в 1978г.

В 80-е годы ведущим специалистом по философии ИИ стал Джон Сёрл. Но и он критиковал как сам ИИ, так и его когнитивные способности.

Вышедшая в 1989г. книга Роджера Пенроуза «Новый ум короля. О компьютерах, мышлении и законах физики» [13], наделала много шума.

«В своей книге Пенроуз утверждает, что человеческое сознание не является алгоритмическим, и в силу этого не может быть смоделировано с помощью обычного компьютера типа машины Тьюринга. По мнению Пенроуза, для понимания природы человеческого сознания важную роль должен сыграть аппарат квантовой механики, в частности, редукция фон Неймана. Пенроуз считает, что необходимо разработать новую теорию, которая будет включать в себя «объективную редукцию волновых функций».
<https://ru.wikipedia.org/?curid=6209191&oldid=124185712>

Другими словами Пенроуз говорит, что мозг представляет собой объект, в котором действует «редукция фон Неймана» - «...мгновенное изменение описания квантового состояния (волновой функции) объекта, ...а из мгновенности

изменения формально следует распространение взаимодействий быстрее скорости света,... Несмотря на мгновенность действия, при редукции принцип причинности не нарушается и информация не передаётся.»
<https://ru.wikipedia.org/?curid=26482&oldid=125672559>

При этом Пенроуз считает, «что квантовое состояние находится в суперпозиции до тех пор, пока кривизна пространства-времени не достигнет значительного уровня. ...Это альтернатива копенгагенской интерпретации, в которой коллапс волновой функции происходит в момент наблюдения, и многомировой интерпретации, которая утверждает, что альтернативные исходы также существуют, но ненаблюдаемы из-за декогеренции. ...Эта интерпретация является теорией объективного коллапса. В этих теориях волновая функция испытывает коллапс как объективный процесс, при этом наблюдатели не играют никакой особой роли. ... волновая функция не может находиться в суперпозиции за пределами определенной разности энергий между квантовыми состояниями. Эту разницу он называет "уровнем одного гравитона" (по всей видимости имеется в виду планковская масса). Несмотря на трудности точного определения, он предполагает, что базовые состояния, в которых происходит коллапс, должны описываться стационарными решениями уравнения Шредингера-Ньютона.»
<https://ru.wikipedia.org/?curid=8399347&oldid=109731929>

Правда при этом: «Пенроуз отмечает, что его идеи о природе сознания являются спекулятивными, и его тезисы вызывают критику со стороны экспертов в области философии, компьютерных наук и робототехники».
<https://ru.wikipedia.org/?curid=6209191&oldid=124185712>

Т.е. его высказывания являются отвлеченными рассуждениями и не имеют под собой подтверждения опытом. Но, согласитесь, красиво!

Его версия сразу выводит человеческий мозг за пределы не только тела, но и нашего мира. Всё, что в нем происходит нам недоступно, потому, что работает «многомировая интерпретация», «...которая предполагает существование, в некотором смысле, «параллельных вселенных», в каждой из которых действуют одни и те же законы природы и которым свойственны одни и те же мировые постоянные, но которые находятся в различных состояниях. Исходная формулировка принадлежит Хью Эверетту (1957 год).»
<https://ru.wikipedia.org/?curid=151275&oldid=126611106>

Круто? Потому, что надо соответствовать! ММИ является одной из многих многомировых гипотез в физике и философии. На сегодняшний день она является одной из ведущих интерпретаций, наряду с копенгагенской интерпретацией и интерпретацией согласованных хронологий.

Добавляет ли это что-то к пониманию хоть человеческого, хоть искусственного разума и сознания?

Вряд ли... Скорее уводит от истины в область фантастических теорий.

Далее можно привести пример книги философа Д.Чалмерса «Сознающий ум. В поисках фундаментальной теории.» [14], вышедшей в 1996г. Направление то же, которое задано Д.Сёрлом, а до него С.Лемом. Всестороннее рассмотрение ИИ как конкурента человеческого интеллекта, с философских позиций, естественно.

Может ли ИИ противостоять человеку, есть ли у ИИ воля и где сознание, насколько он «дружественный» человеку, как ИИ должен принимать «справедливые» решения и следовать моральным и этическим нормам человека?

Я не понимаю, как можно задавать такие вопросы, после полувекового развития ИИ на основе математических методов его построения?

Мы кому эти вопросы задаем? Математикам, которые нашли методы оптимизации сетевых решений, или программистам, которые их реализовали?

Ни те, ни другие ни о чем подобном никогда не думали при реализации поставленных им задач.

Тем не менее вопросы услышаны и последовало организационное решение:

«Продвижение Дружественного ИИ является одной из основных задач Singularity Institute for Artificial Intelligence, равно как и получение финансирования этого проекта и в конечном счёте создание «зародыша ИИ», в систему мотивации которого будет изначально встроена забота о человечестве. Несколько известных футурологов высказались в поддержку теории Дружественного ИИ, включая писателя и изобретателя Рея Курцвейла, специалиста по продлению жизни Обри ди Грея и сооснователя World Transhumanist Association Ника Бострома.

11 декабря 2015 года Илон Маск, Сэм Альтман и другие основали компанию OpenAI, целью которой является содействие созданию открытого ДИИ.

...Наиболее заметным критиком теории Дружественного ИИ является Билл Хиббард, автор книги «Сверхинтеллектуальные машины», который считает эту теорию неполной. Хиббард пишет, что для создания ДИИ требуется большее политическое обсуждение вопросов создания ИИ и ИИ-морали.» <https://ru.wikipedia.org/?curid=2364621&oldid=128737741>

Таким образом, общефилософская концепция не изменилась.

Она всё так же направлена на моральную и этическую стороны ИИ, не рассматривая его технические характеристики. При этом все философы ИИ критикуют «разумность» любого интеллекта, включая и человеческий, которая с математической точки зрения ничем не обоснована.

И в общем, надо сказать, критика не лишена оснований. Действительно, сегодня хороший алгоритм «когнитивного моделирования» позволяет машине пройти как тест Тьюринга, так и многие другие тесты на разумность [6]. Единственный путь сближения указан как разработка «дружественного» интеллекта.

С технической стороны, всё ровно наоборот.

ИИ самоотстранился от естественного интеллекта и взял курс на математические пути развития, исключая возможность соединения человеческого и машинного интеллекта в один комплекс. Как показывают практические результаты развития, западные ученые идут к созданию суперинтеллекта на основе суперкомпьютера, минуя процесс систематизации всех возможных интеллектов.

Картина мира.

Все философы, пишущие о проблемах ИИ, почти автоматически его «очеловечивают», наделяя свойствами, характерными для человеческого интеллекта. Вводят его в состав человеческого общества и рассматривают его в отношениях с человеком и обществом.

Потом такой «очеловеченный» интеллект начинают рассматривать в условиях копирования им вполне «человеческих действий» по отношению к человеку. Как такой ИИ а) противодействует человеку; б) нейтрален к человеку; в) дружелюбен человеку. Не знаю, насколько это правомерно в применении к ИИ, но результат в этом случае всегда один и тот же. Из машинного интеллекта или разума получается человек.

Видимо философы сначала представляют себе ну, что что-то такое, может быть «машинный мозг на ножках», я не знаю, но как-то так. В какой-то момент этот «мозг на ножках» оказывается в кругу людей, живет с ними, ума набирается, а потом вдруг начинает мыслить как они. Это значит, между прочим, что он учится врать, скрывать свои намерения, строить ответные реакции на действия людей в соответствии с их нормами морали и этики так, как он это понял при общении. И вот тут философы, пишущие об ИИ вдруг начинают бояться того, что может сделать этот самый машинный «мозг на ножках», если вдруг и если что....

Или идут более древним путем ...

Ранее человек переносил страхи и опасения на всё, что видел вокруг. На гору, на ветер, на Луну и Солнце, и т.д. одухотворяя их, делая их божественными созданиями имеющими Дух. Вот эти духи и становились тем активным началом, с которым можно было разговаривать на человеческом языке. Дух горы, это и есть «живая гора», дух ветра это и есть ветер, к которому обращается человек.

При этом, дух отдельно, гора – отдельно. Гора существует самостоятельно, на неё можно взбираться и что-то там делать, при этом есть и дух горы, который может как мешать человеку в его делах, так и помогать. Духа можно вызвать на разговор, что-то объяснить ему и попросить у него помощи.

Если так, то у нас есть ещё и «дух машины» - интеллект или разум, к которому обращаются философы, пытаясь постичь «человекоподобную» машину, если она вдруг появится. Потому, что «дух машины», это же и есть та машина, которую они постигают через «душу», как какое-то подобие человека.

Это и называется «очеловечиванием» машины. Если говорить на языке философии, это переход от принципа антропоцентризма к антропоморфизму. Перенос человеческих чувств и отношений на неодушевленные предметы.

В данном случае, на машину.

Проблема тут возникает глобальная. Философская.

Что главнее – душа или тело?

Для философии тут ответ очевиден – душа.

В оценках ИИ западная наука сегодня склоняется к тому же варианту возможного ответа. В этом смысле, для западной науки медленно, но верно, *ИИ превращается в «духа машины», виртуальную самостоятельную сущность.*

Пока просто отметим этот факт.

Если сравнивать советский проект «Искинт» и западный ИИ, можно четко увидеть почти полярность подхода к построению искусственного интеллекта для машины. Это же полярность «материальное – виртуальное» для реализации искусственного интеллекта.

С одной стороны, появление математических разработок нейросетей практически решило эту проблему очень оригинально. «Вместилище» «духа» программного ИИ, современная *нейронная сеть*, имеет в основном виртуальное существование и работает в режиме симуляции. На мощном суперкомпьютере стандартной архитектуры и высокой производительности.

Таким образом, вся система «программа+нейросеть» оказалась полностью виртуальной и вполне работает на любом компьютере стандартной конфигурации и архитектуры.

В виртуальном пространстве виртуальной нейронной сети существует «дух машины», программный ИИ, который мы считаем реальным.

Это так по-человечески...

Сегодня никто не представляет себе ни «гиромат», ни «машинное сознание», ни «модель мира», ни даже, что такое ИИ!

Но с другой стороны есть проблема...

Представлении ИИ в виде «духа машины», это начало идеализма для ИИ.

Второй вариант – «мозг на ножках». Это материализация ИИ в виде объекта, который мы должны рассматривать «целое», включающее как сам ИИ, так и место его расположения.

Если модели ИИ соединить с картиной мира на рис.3., взятой в <https://dzen.ru/a/X2N3uBdPmg5tOC18>, то что мы увидим?

Советский Искинт и западный ИИ оказываются крайними точками зоны расположения всех возможных систем, содержащих элементы интеллекта и разума. На рис.3. это фиолетовая горизонтальная прямая.

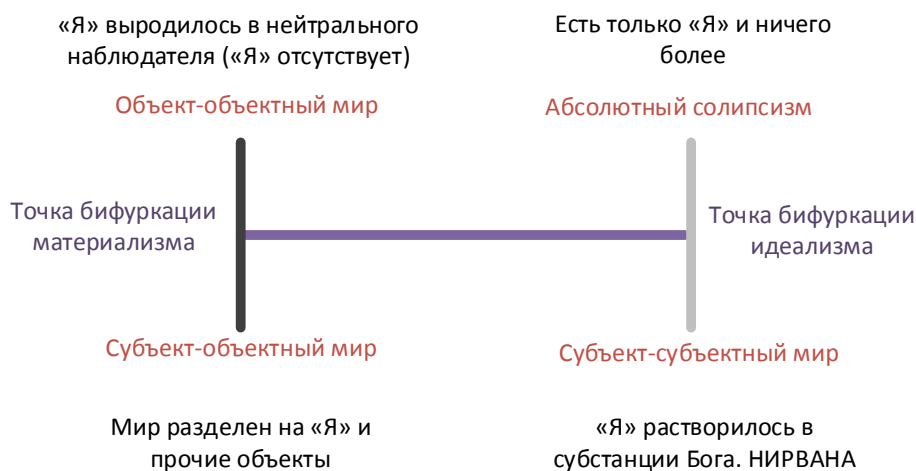


Рис. 3. Картина мира в зависимости от статуса наблюдателя.

При этом должен быть субъект или объект рассмотрения.

И возникают главные вопросы:

- Интеллект, это что - объект или субъект?
- Интеллект, он какой - материальный или виртуальный?

В зависимости от полученных сейчас ответов мы сможем продолжить рассмотрение далее. Но философия таких вопросов не ставила, а наука ответов не давала.

И всё же, советский проект Искинта идет в сторону «субъект-объектного мира». В рамках проекта Искинт – субъект. Тогда вся философская платформа человека должна как-то распространяться и на этого вполне материального субъекта – «машину, обладающую искусственным интеллектом».

Для этой машины и название нашлось – кибер. В конечном варианте предусматривается материальная субъектность системы Искинт, как кибера [18], обладающего каким-то интеллектом.

Д.А.Поспелов указывает на этот отличительный момент:

«Единственной возможностью расширения способностей машины к решению творческих задач, адаптации в неисследованной среде, самоорганизации и т. д. является создание внутри нее семиотической системы ее внешнего мира (например, мира задач). Окружающий машину мир обладает определенной структурой, представляет из себя совокупность объектов, связанных между собой сложными системами связей. Этот мир

может обладать динамикой, в нем могут реализоваться различные детерминированные или вероятностные закономерности. Сложность отображения этого мира в памяти машины связана с необходимостью отображения его структуры, его динамики и законов взаимодействия мира с машиной. При этом предметом дальнейшего исследования в машине является каким-то образом сформированная модель мира, адекватная ему с точки зрения системно-структурного подхода.» [5]

Вот эта внутренняя «модель мира» и стала основной проблемой построения Искинта и приближении его возможностей к человеческому разуму.

Западное направление развития ИИ изначально не предполагало субъективизации, да и материализации тоже. Программные алгоритмы постепенно увеличивались и усложнялись и в объемах и в сложности. От создания экспертной системы на одном процессоре, до программирования многопроцессорных комплексов нейропроцессоров, а потом и нейросетей.

Но и на новом уровне ничего не изменилось.

Нейросеть виртуальная, схема и алгоритмы логической обработки информации не предполагают использования другой реальной сети кроме Интернета, в формате обмена информацией с отдельными пользователями.

Полученный таким образом программный комплекс ИИ не очень привязан к конкретному компьютеру и не составляет с ним единого Целого. Здесь Целым является только сам программный ИИ, более виртуальный, чем материальный. Даже если ввести в состав этого комплекса средства активного взаимодействия с человеком, они не сделают ИИ самостоятельным. Он всегда будет идти ведомым в этой связке.

Западный программный ИИ, как истинный «дух машины», как «раб лампы», джин из сказок про Алладина, на рис.3. находится на прямой «идеализма» и может пойти в любую сторону.

Например в сторону «субъект-субъектного мира» с растворением программного ИИ в человеческом программировании. Тогда никакого самостоятельного ИИ не может быть, а вместе с этим невозможны и никакие «машинные войны».

Развитие ИИ может пойти и в сторону солипсизма. Здесь вариант «солипсизма», это самостоятельность на уровне сложного вируса, бродящего в сети интернета. Конечно, такой виртуальный монстр долго не проживет. Каналы передачи информации уже хорошо контролируются со всех сторон.

Но наделать бед он точно может.

«Раб» или «друг»?

Когда снова и снова перечитываешь материалы из мировой истории развития ИИ, невольно напрашиваются и политические аналогии.

Ну, смотрите сами...

Западная школа ИИ изначально ограничила «машинный разум» до «интеллекта», как «вычислительной составляющей способности достигать целей в мире». В этом отношении, «робот», как «раб», самое точное понимание для обозначения цели западного варианта движения к построению ИИ. Запад так и идет в этом направлении. В сторону построения себе новых «слуг», облегчающих жизнь их «творца». А как же? Человек управляет всем.

Западный вариант программного интеллекта сегодня стремится удержать свойство достигнутой виртуальности, даже в условиях усиливающейся активной

самостоятельности, которую разработчикам приходится вводить в его алгоритмы работы. При этом все понимают, универсальность виртуального интеллекта вступает в противодействие усилению его универсальной самостоятельности. Мы прямо видим, как «легким движением руки брюки превращаются..., брюки превращаются...». Простите, техническая неувязка.

С другой стороны, советская наука когда-то сразу определила Искинт материальным объектом сложного состава «программа+машина». Ну если несколько фривольно, тот самый «мозг на ножках».

Советская школа Искинта изначально начала строить себе «друга», расширяя понимание «машинного разума» до философских пределов человеческого. Обдумывались варианты «думающих» машин, обладающих «сознанием», с использованием «самообучающегося центра принятия управляющих решений» - гиromата. Именно об этом говорят все работы Д.А.Поспелова, его коллег и последователей.

В советской фантастике 50-60-х замелькал свой аналог «робота» - «кибер», «самостоятельный» автомат с каким-то уровнем субъектности, в отличие от «робота» [15]. Он так и рассматривался, как «механический» друг, помощник, которого надо уважать. На такой искусственный интеллект была нацелена советская наука.

Кибернетический подход к пониманию машинного интеллекта, как необходимой части машины определил всё дальнейшие шаги развития машинного Искинта в СССР. Искинт стал проектом реализации кибера из советских фантастических романов, машиной, обладающей каким-то уровнем машинного интеллекта, какой-то самостоятельностью и даже субъектностью в принятии исполнительных решений и их реализации.

Для него и разрабатывался гиromат, как система самостоятельного управления, какой-то блок машинного сознания и всё это объединялось в интеллектуальную систему логической обработки информации с учетом этических и моральных норм, принятых в человеческом обществе.

Тут возникает сложная дилемма в реализации ИИ.

Или это алгоритм с элементами «вирусной» активности, программы виртуального субъекта неизвестного уровня самостоятельности... и естественное продолжение взятого курса... в тупик глобализации проблемы.

Или это привязка к конкретному «железу», где работает программа ИИ. А это уже равносильно отказу от виртуальности ИИ, как «духа машины» и постепенный вынужденный переход к его материализации в системе Искинта, как «программа+машина» и движение к дальнейшей его индивидуализации в виде «мозга на ножках».

Правда, для такого перехода у науки и сегодня пока ничего нет. Ни теории, ни технологий, ни, самое главное, желания. Потому, что это, в первую очередь, потеря объемов отлаженного производства и соответственно, прибыли от продаж уже привычного программного ИИ. Зачем резать курицу, несущую золотые яйца?

А время идет. Прогресс не остановить. Не сделаешь ты, сделают другие и тогда ты уйдешь в тень. Чтобы удержаться наверху надо что-то делать.

Какой подход лучше?

Тут простого ответа нет.

Можем мы сегодня сказать, чем различаются модели «Друг» и «Раб» кроме их названий?

Реально – вряд ли. Причина проста.

По начальным установкам Искинта, чтобы машина стала умной надо...

1. Создать машине «модель мира». Как у человека. Чтобы машина соотносила свои решения с тем, что входит в её модель мира, применяла лишь те решения, которые соответствуют этой модели, остальные отбрасывала, как неэтичные и аморальные.
2. Создать машине «сознание», как место, где и происходит процесс решения. Как мыслительный процесс у человека.

Часть ученых полагают, что если у машины есть «сознание», то должны быть и интенции, это «намерение; стремление» — направленность сознания, мышления на какой-либо предмет. На те самые цели, которые надо достигать...

Советский проект Искинта был ориентирован на кибернетический подход в реализации искусственного интеллекта, как основы самоуправления. На это был нацелен «гиромат» Д.А.Поспелова [5,16]. Эта цель была четко артикулирована.

Сегодня часть ученых сопротивляется этому, говоря, что никаких интенций у машины быть не может [17]: *«У машины не может быть интенции, не может быть интуиции и инициативы, нет воображения, она не способна быть творцом, но может быть мощным его партнером.»*

Это вариант движения по направлению, взятому западной наукой. Западный проект нацелен только на многоуровневую логическую обработку и рост объемов обработки информации. Здесь ИИ стал вычислительной машиной нового типа. Виртуальной нейросетью.

Советский проект опередил время и был свернут, а западный хоть и был изначально ограничен, продолжил развитие, и привел ИИ вот в эту точку, где он сейчас и пребывает. Развитие вроде есть, но сегодня оно больше «виртуальное», чем реальное. ИИ можно лишь моделировать на искусственных нейросетях, также смоделированных в компьютере. Потому, что все эти «нейросети», ИИ, это всё тот же, один, хоть и большой, компьютер огромной производительности, работающий с виртуальной моделью, а не с реальной нейросетью.

По сути дела, сегодня мировая наука создала Программный ИИ вместо реального и пытается на его основе решать реальные задачи имея в наличии лишь суперкомпьютеры стандартной архитектуры.

Пока это как-то получается.

*г. Волгодонск
май 2023г*

Литература:

1. Никитин А.В., В плену «призрачных истин» // «Академия Тринитаризма», М., Эл № 77-6567, публ.28241, 28.12.2022 <http://www.trinitas.ru/rus/doc/0016/001h/00165190.htm>
2. Никитин А.В. На пути к Машинному Разуму. Круг третий. (Части 1,2) // «Академия Тринитаризма», М., Эл № 77-6567, публ.12887, 31.01.2006 <http://www.trinitas.ru/rus/doc/0023/001a/00230029.htm>
3. Дрейфус Х. ЧЕГО НЕ МОГУТ ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ 1978 <https://sheba.spb.ru/za/chegonemozhet-evm-1978.htm>
4. Системные исследования. Ежегодник. 1969 М.: Наука, 1969.- 204 с. <http://www.sci.aha.ru/ots/doc/sys1969.pdf>
5. Поспелов Д.А. Сознание, самосознание и вычислительные машины// Системные исследования. - М.: Наука, 1969. http://www.raai.org/about/persons/pospelov/pages/SR1_Pospelov_1969.pdf

6. Станислав Лем. Магелланово Облако
<https://www.booksite.ru/fulltext/o/001/005/167/049.htm>
7. Советская ассоциация искусственного интеллекта надежды и стремления
[http://www.swsys.ru/index.php?page=article&id=1399&lang=,](http://www.swsys.ru/index.php?page=article&id=1399&lang=)
8. Тарасов В.Б. Искусственный интеллект: от психоники и прикладной семиотики к синергетическому подходу. О научном наследии профессора Д.А. Поспелова <https://computer-museum.ru/articles/materialy-mezhdunarodnoy-konferentsii-sorucum-2020/5167/>
9. Новости Искусственного интеллекта". №1 1991 г.
<https://cloud.mail.ru/public/GSuW/k6b8iEKqV>
10. Психоника <http://www.geotar.com/geota/psychonica/psyo1.pdf>
11. В. А. Лефевр Конфликтующие структуры
https://gtmarket.ru/files/book/7307/Conflicting_Structures.pdf
12. А. Г. Чачко Искусственный разум 1978 Серия «Эврика»
<https://sheba.spb.ru/za/evrika-isk-razum-1978.htm>
13. Роджер Пенроуз, Новый ум короля. О компьютерах, мышлении и законах физики.
https://royallib.com/read/penrouz_rodger/noviy_um_korolya_o_kompyuterah_mishlenii_i_zakonah_fiziki.html#0
14. Дэвид Джон Чалмерс. Сознательный ум. В поисках фундаментальной теории. 1. Возможен ли искусственный интеллект?
<https://fil.wikireading.ru/17241>
15. Никитин А.В., Когда появятся киберы? // «Академия Тринитаризма», М., Эл № 77-6567, публ.25481, 04.06.2019
<http://www.trinitas.ru/rus/doc/0016/001g/00164050.htm>
16. Поспелов Д.А. Фантазия или наука
https://www.phantastike.com/superlearning/fantasy_science/djvu/
17. Ваганов А., Интервью с В.Финном: 13 признаков искусственного интеллекта https://www.ng.ru/nauka/2018-06-27/9_7253_intelligence.html
18. Бостром Н., Искусственный-интеллект. Этапы. Угрозы. Стратегии
<https://avmim.com/wp-content/uploads/2018/11/Бостром-Н.-Искусственный-интеллект.-Этапы.-Угрозы.-Стратегии-2014.pdf>