– Т.5. От сверхдержавы СССР к России постсоветской (1945-2000 гг.) – Самара, 2005.

5. Горожанин А.В. и др. Русь - Россия - СССР - Россия: этапы большого пути: Учебное пособие для вузов по курсу "Отечественная история" / под общей редакцией В.В. Рыбникова. – Т.1. От Руси изначальной к году 1917: "Из мрачной глубины веков ты поднималась исполином...". – Самара, 2004.

6. Горожанин А.В. и др. Русь - Россия - СССР - Россия: этапы большого пути: Учебное пособие для вузов по курсу "Отечественная история" / под общей редакцией В.В. Рыбникова. – Т.2. По исторической спирали взлетов и падений бурного XX века: «в года разлук, в года сражений, когда свинцовые дожди...». – Самара, 2006.

ИСКУССТВЕННЫЙ – ЭТО ЕСТЕСТВЕННО, ИЛИ ПРОТИВ АРТИФОБИИ

Ашкинази Л.А.
Кандидат физико-математических наук, преподаватель физики ФМШ МИЭМ, редактор журнала «Химия и жизнь», г. Москва

Страх перед искусственным интеллектом,
артифобия (от artifical intelligence) – это разновидность ксенофобии.
Но фобии и подозрения насчет коварства ИИ – инфантильны.
Человек проецирует на ИИ свой врожденный,
биологически мотивированный эгоизм.
Михаил Эпштейн

Ситуация

Мы видим рост интереса к тому, что принято называть искусственным интеллектом (далее – ИИ; естественный обозначим ЕИ). Причин две, первая – достижение определенного уровня развития области, выразившегося в AlphaGo, Watson и особенно в GPT-4. Вторая – человеческая психология: результаты работы этих программ похожи на результат деятельности человека, достаточно умного для того, чтобы оказаться и полезным, и опасным.

Мы видим высокую концентрацию эмоций и поляризацию, преобладание либо восторга, либо страха. Причина страха – психологическая: человеку и хочется, чтобы эта могучая сила всё за него делала, и чтобы беспрекословно подчинялась. Ощущая, что это вряд ли совместимо, человек начинает бояться. При этом он приписывает ИИ тенденцию к совершению человеческих мерзостей, которые видит вокруг. Не понимая, что люди – участники и продукт определённой истории, а у ИИ такой истории нет.

Разумеется, ИИ имеет доступ к этой истории – как через всю имеющуюся информацию, так и через людей – создателей и собеседников. Но это всё-таки не его история, и знание о чём-то влияет слабее, чем участие. Например, люди имеют доступ к истории, однако сам по себе он не сделал людей в целом мерзавцами и негодяями.

Страх пользуется спросом на рынке. Часть прессы паразитирует на психологии, причём именно на страхе. При этом используются умные слова, не имеющие отношения к делу – фазовый переход, огромная скорость, нелинейность, экспонента, особая точка, сингулярность, синергия и т.д. Потом вокруг темы возникают организации, обеспечивающие людям занятость свободного времени и ощущение собственной значимости, потом подгребают паразитирующие на этом хайпе политики. Такой ход событий в последние десятилетия мы уже видели.

Попутно – мне кажется, что мы вообще пренебрегаем влиянием психологии исследователей на эволюцию наук. Например, циклические модели Вселенной могут для кого-то казаться предпочтительнее моделей неограниченного необратимого расширения именно по психологическим причинам (а для кого-то наоборот). И это несмотря на невообразимое соотношение продолжительности нашей жизни и космологических времен. Интересно, имеет ли значение психология для эволюции других областей физики, например, физики элементарных частиц?

Начнем с рассмотрения простых ошибок, которые мы часто делаем при обсуждении проблем ИИ, а потом обсудим более интересное. Данная статья посвящена именно анализу ошибок, которые мы делаем при обсуждении проблем ИИ. Мой позитив был опубликован ранее; его можно увидеть, в частности, тут: [6; 2].

Ссылки на авторитет, которые ничего не стоят

Статьи в естественных науках можно разделить на две группы – оригинальные статьи и обзоры. В оригинальных статьях ссылки выполняют несколько функций, обычно это указание на источник информации. Иногда сообщается, кто ещё занимался этой задачей – даже если их результаты в самой статье не используются. Что касается обзоров, то их пишут специалисты, работающие в соответствующей области, они иногда приводят в обзоре и свои новые результаты. Так что граница между оригинальными публикациями и обзорами не абсолютна. Но «соседний» материал занимает в статьях мало места, и странно выглядела бы сегодня ссылка на Ньютона в статье о движении спутника, или на Максвелла в статье о работе гиротрона.

В гуманитарных областях это не так, ссылки часто используются, чтобы сказать – A, B и C думали так же, как я, и великие D и F – тоже. Понимать это надо так – мне в институте разжевали и положили в ротик про C и D, про F мне кто-то рассказал, нечто подобное у A я отыскал сам, правда, слабо подобное, но кто ж будет проверять, а про B я не проверял, но он писал вообще обо всём на свете, наверняка и об этом тоже.

В любой гуманитарной области со временем достигается такой уровень развития, что для любого мнения, поддержанного авторитетами, можно найти противоположное мнение, тоже кем-то поддержанное. Поэтому ссылки на то, что кто-то что-то сказал, вообще ничего не доказывают. Кстати, слова «поддержанный» и «подержанный» тревожно близки.

Немотивированные утверждения

Употребление слов «невозможно», «не может быть», «невозможно себе представить», «никогда» и т. п. означает только одно – пишущий испуган и хочет, чтобы читатель разделил с ним груз испуга. Груминг – это прекрасно! Особенно характерно неоднократное употребление этих слов в одном тексте. В шоу-бизнесе этот прием стандартен, соответствующий текст называется «припев». Вот что пишут профессионалы:

«Припев является повторяющейся и самой «цепляющей» частью песни, которой в первую очередь хочется подпевать. С точки зрения слов песни, задача припева – выразить в целом идею всей песни и выразить её название. Припев должен быть таким, чтобы его можно было быстро запомнить. В припеве не нужно вносить какую-то новую информацию, новые детали, для этого существуют куплеты».

Разница между шоу-гипнотизёрами и нами, пишущими про ИИ, в том, что они не испуганы тем, о чем поют, а с нами такое иногда бывает.

Откуда мы и ИИ получаем информацию

На психологию, и в том числе на интеллект человека влияет получаемая информация. Она поступает при самостоятельном изучения мира, от других людей, из источников информации (книг, баз данных, кино и т. д.) Иногда утверждается, что ИИ в принципе лишён части этих каналов, но это не так. ИИ может общаться и с другими ИИ, и с людьми. Соответственно, для него нет принципиальных препятствий для общения с обществом, хоть человеческим, хоть обществом ИИ. Общение с «Кьюриосити» и «Персеверансом» будет изучением окружающего мира, а если обеспечить связь между ИИ и андроидом, то он сможет изучать и своё тело – как младенец, тянущий в рот большой палец правой ноги. Не приходило ли это в голову многочисленным уже существующим андроидам? К некоторым телескопам и видеокамерам есть доступ через интернет, и ещё есть «интернет вещей» – всё это внешний мир, причём более доступный для ИИ, чем для нас. Причём лишение человека большинства источников информации не исключает развития интеллекта [9].

Создаёт ли ИИ нечто новое

Иногда утверждается, что ИИ не создаёт что-либо новое. Интересно, считают ли те, кто так говорит, чем-то новым партию в го или шахматы, выигранную у чемпиона мира? Что думают по этому поводу люди, серьёзно разбирающиеся в великих играх и соответствующих программах? GPT-4

использует как источник информации интернет, и при обучении пользуется отзывами людей. Ровно так же учится любой человек – используя учебники и преподавателей (не у всех есть лабы). Поэтому, если не создаёт новое ИИ, то возникают сомнения в новизне созданного ЕИ, человеком.

Сравнение с человеком и с идеалом

При обсуждении проблем ИИ мы часто прибегаем к сравнению ИИ с человеком. Эта традиция восходит к Тьюрингу, причем «тестом Тьюринга» мы часто называем не то, что именно он предложил; и вообще там есть много весьма интересных вариантов. Хорошее обсуждение тут [22].

Но дело не в том, как называть, дело в том, что люди разные – с кем из них надо сравнивать? У психологов есть несколько тестов интеллекта, которые выдают большую и содержательную картину – почему бы не прогонять ИИ по ним?

Но всё это – сравнение именно с человеком. А может ли существовать нечеловеческий интеллект? Да, у обезьян интеллект явно есть, но мы воспринимаем его, сравнивая с человеческим. А существует ли не сверхчеловеческий, и не недочеловеческий, а просто «не», просто другой? Интеллект – это часть психологии, задача «придумать психологию» не кажется безнадежной [7], как и соседняя задача — придумать социологию [1].

Рассуждая формально, придумать другую психологию и социологию было бы проще всего, поняв, частью чего они являются, и посмотрев, что лежит рядом с ними. Другой путь – разделив их на части, и либо соединив эти же части иначе, либо изменив их список.

Сравнивая ИИ с человеком, иногда переходят (может быть, не заметив этого) от сравнения с человеком к сравнению с идеалом. Эта ситуация имеет два уровня – когда идеал существует хотя бы в принципе, и когда он вообще не существует и не может существовать. Пример первой ситуации – отношение к автомобилю, управляемому ИИ. Такой автомобиль радикально безопаснее обычного, но некоторые готовы примириться с ним лишь при абсолютной гарантии безопасности. Которой быть не может, хотя бы потому, что человек может броситься под колёса или отправить туда ближнего своего. Чего не сделаешь, чтобы доказать, что робоавтомобиль опасен.

Конструируя ситуации, которые должны по их замыслу испугать, авторы иногда нарушают логику. Например, пишут: «Если полицейские отряды, состоящие из людей, могут отказаться выполнять какие-то драконовские директивы (например, убивать всех представителей определенной демографической группы), то такая автоматизированная система не поддастся никаким угрызениям по поводу проведения в жизнь любых капризов вышестоящих человеческих существ». Имея в виду попугать машинами, автор проговорился – «капризы человеческих существ». Кроме того, мы видим – есть миллионы людей, которые прекрасно «не поддаются угрызениям».

Вот ещё пример ситуации, предложенной мне одним из оппонентов (текста нет в интернете и автора не называю). «Человек садится в своё беспилотное авто, говорит адрес, машина блокирует ручное управление и двери и привозит человека в полицейский участок. Почему? Потому, что адрес – вблизи места сбора митинга оппозиции. Или ИИ, зная, что по адресу будет такой митинг и, зная, чем он может закончиться, заботится о безопасности хозяина (в строгом соответствии с законами роботехники Азимова) и в полицию не повезёт, а просто увезёт хозяина из города и не выпустит, пока митинг не закончится. Такая вот этика и гуманизм ИИ».

Автор ситуации не обратил внимания на более простые решения. Не тратить бензин, а просто запереть и стоять на месте, или запереть в их машинах всех, кто собирался приехать на митинг. Или – беречь так беречь! – запереть всех в домах. Вот в мире каких кошмарных фантазий живут некоторые; отнеситесь к ним с пониманием и сочувствием. Но обратите внимание – эти кошмарики сконструировали умные люди и серьёзные специалисты, причём сконструировали именно на базе знания людей, знания нашей человеческой истории.

Вот ещё одна страшилка «от оппонента» (текста нет в интернете и автора не называю).

«Родителям ребенка с диагнозом «дебилия» предлагают вживить что-то ему в мозг. Они согласны. Ребенок становится первым учеником, победителем олимпиад, пишет стихи и рисует картины, сочиняет музыку, играет на инструментах, знает несколько языков, поступает в университет. Родители в восторге, но это – не их ребенок, тело их ребенка стало аватаром ИИ. Чип в мозгу – только средство связи. А сам мозг только управляет физиологическими функциями».

Ненадёжны могут быть не только беспилотные автомобили и промышленные роботы. Любая программа может выдать неправильные цифры – из-за ошибок, сделанных её автором, ошибок других программ и сбоев железа. Мы привыкли к этому, и как-то научились с этим умеренно справляться, хотя самолеты иногда перепутывают верх и низ. Но перепутывает и человек – была версия, что сбить хотели своего, чтобы получить casus belli, а по криворукости грохнули чужого. Но программы, на коробке с которой написано ИИ, человек боится больше, и поэтому предъявляет к ней больше требований. И совершенно зря – потому, что он сам ошибается чаще этой программы.

Алгоритм преобразования алгоритма преобразования алгоритма

Иногда в качестве принципиального отличия ИИ от ЕИ говорят, что ИИ «действует по алгоритму». Но нейрон тоже действует по алгоритму, а то, что нейронов у человека много, не делает их менее алгоритмичными. Если мы всё-таки свяжем интеллект с количеством нейронов, нам придётся отвечать на вопрос – где граница? Если мы осмелимся ответить, компьютерщики предъявят систему, которая не слабее. Если нам скажут, что

в компьютере алгоритм не меняется, а в человеке меняется, то мы ответим, что алгоритм в компьютере может изменяться, причём само изменение может быть изменяемым и так далее сколько угодно раз. То есть имеются в виду не последовательные изменения одного алгоритма, а вложенные, то есть на каждом уровне изменяется нижележащий алгоритм.

На аргумент о непредсказуемости происходящего в мозгу человека мы спросим, как вы отличаете непредсказуемость от простого человеческого незнания или от случайных процессов, а в ответ увидим закатывание глаз и услышим произнесение с придыханием «голография» и «квантовые процессы».

Тут ощущается горячее дыхание серьёзной проблемы. Проблема вложенных алгоритмов – это часть «проблемы возникновения нового». Когда при усложнении структуры и алгоритмов возникает «новое» и почему? Вот вариант простого и конструктивного ответа – когда мы видим изменение наблюдаемых нами признаков. Когда возникает речь, использование орудий, хранение орудий, изготовление орудий, передача навыков, рисунки на стенах пещер… А на вопрос «почему» у нас нет ответа. Когда мы сможем загрузить в машину времени установку фМРТ, то сможем узнать, какие структуры в мозге работали, когда наш предок делал что-то новое. Такое исследование интересно проделать и в современности, и есть попытки [21].

ИИ и этика

От ИИ могут требовать решения этических задач, которые не имеют общего решения. Классический пример такой задачи – «проблема вагонетки», о ней лучше всего прочесть здесь [15].

Решения у подобных задач нет не потому, что мы его не знаем, а потому, что мы знаем их много; причём разные «мы» знают разные наборы решений. Как на персональном уровне, так и на уровне групп (половых, возрастных, населения стран, религиозных групп и т. д.). Содержательный анализ тут: [13].

Более того – индивидуальные мнения могут в некоторых случаях зависеть от многих факторов, о которых мы даже не подозреваем. А на групповом уровне – от деятельности пропаганды, которая, как оказалось, вполне может натравить один народ на другой. Я имею в виду гитлеровскую Германию.

Человечество имеет опыт таких ситуаций – сначала это межплеменные и религиозные войны, а потом голосования и прочие признаки цивилизации. Хочется надеяться, что опыт цивилизованного решения этических проблем будет людьми расширен. Попутно – чтобы было меньше обиженных, можно ввести компенсации проигравшим за счёт выигравших. Причём до голосования выяснять, какую компенсацию победителей к побеждённым все в среднем сочтут правильной (например, дневной или недельный доход). Социологи и психологи вполне могут решить эту задачу.

Проблема «ИИ и этика» может осложняться тем, что для ИИ, как кажется некоторым людям, надо разработать формальное решение, а это воспринимается как посягательства на человеческие прерогативы. Однако мы знаем, что ЕИ может обучаться той или иной этике (в том числе и этике, предписывающей массовые убийства в отношении другой этнической группы). Наверное, так же этике может обучиться ИИ. Причем в этом случае отпадает возражение о формализации, потому что выработка формализации вообще не имеет места.

Но возникает другая проблема – мы видим, что люди в смысле «этики» очень разные. Можно осторожно предположить, что и ИИ могут в этом смысле оказаться разными. Будем надеяться, что лучшие айтишники, программеры и собеседники окажутся на светлой стороне. И что мы сами будем именно на ней. Попутно заметим, что с тем, как решает этические проблемы человек, тоже не всё ясно – и может быть, погружение в проблему ИИ позволит (как побочный результат) лучше понять ЕИ. Различия в ИИ будут возникать, как и у ЕИ, в процессе воспитания, взросления. В том числе и непреднамеренно, в результате общения с разными ЕИ. Станислав Лем – рассказ «Ананке» [12].

ИИ, ЕИ и какой-то ещё

Многие проблемы ИИ опираются – так уж мы привыкли – на дилемму «ИИ или ЕИ». Эта проблема не в природе вне нас, а у нас в голове – человеку проще делить все объекты на две кучки. Любой «естественный» интеллект искусственен, потому что он создан влиянием окружения, а не вырос сам по себе. А любой «искусственный» естественен, потому что он создан человеком, частью природы, то есть природой посредством человека. Это, конечно, трюизм, но что вы скажете о композитном собеседнике в тесте Тьюринга, который чередует реплики человека и машины, или смешивает их в любом отношении? Или хитрее – один интеллект предлагает несколько вариантов реплики, а другой выбирает, и далее по очереди? Вариантов организации совместной деятельности немеряно.

Для неё уже применяют термин «коллаборативный», а применение термина – признак определённого уровня развития. В какой момент при встраивании чипов в мозг мы скажем – вот теперь он искусственный? Мы можем попытаться уклониться от этой задачи, введя термин «композиционный интеллект», но из этого ничего не получится, просто нам придется рассматривать не одну границу, а две. Или вводить понятие «степени искусственности» [4].

Так что при любых рассуждениях на тему ИИ надо помнить и учитывать, что между Е и И еще много промежуточного. Даже просто букв – три.

Что касается совместной деятельности, то тривиальных примеров много. Вот один, пока вроде бы не реализованный, но лично мне дорогой и близкий. Человеческая педагогика может быть применена к ИИ, а принципы

обучения нейронных сетей могут проникнуть в человеческую педагогику. Элемент такого подхода в педагогике уже был! В школе, которую я окончил, в начале изучения теории функций школьникам предлагалось самим разработать элементарную классификацию функций, располагая списком функций, но без деления их на классы [3].

В мире ИИ это называется «обучение без учителя» – название неправильное, но устоявшееся.

И вообще – мы всё время обсуждаем, как ЕИ создает ИИ, не замечая, что ИИ уже начал создавать ЕИ.

Постановка целей

Одним из аргументов за принципиальную ограниченность ИИ является отсутствие у него способности к постановке целей и наличие такой способности у ЕИ. Посмотрим, действительно ли есть это различие, а если есть – то как и почему оно возникло, и при каких условиях оно исчезнет. На уровне конкретных действий различий нет – любой выбор ветви в программе можно интерпретировать, как выбор цели. На это можно возразить, что в программе эти выборы прописаны изначально. Но и в человеке выборы прописаны – его внутренним содержанием и окружающим миром. Из этого не следует полная детерминированность человеческого поведения, потому что и внутри человека, и вокруг него есть случайности. Но это же относится и к ИИ – и вокруг него, то есть в мире людей, есть случайности, и в компьютере тоже [6].

Рассматривая выборы, которые делает человек, и спрашивая, почему он сделал конкретный выбор, мы часто видим исходную для этого выбора внутреннюю (психология) и внешнюю (обстоятельства) причину. Разница в том, что программист, как думают непрограммисты, знает про свою программу всё, а человек, обнаружив незнание, закатывает глаза и начинает бормотать о возвышенном. Но профессиональный психолог может существенно расширить наше понимание причин наших собственных выборов.

Идя по человеческому дереву причин и следствий, мы приходим к желанию жить. А оно – простое следствие того, что есть жизнь. Это всего лишь Дарвин – те, у кого не было желания жить, не выжили. Механизмом была и остаётся боль, а у некоторых людей (в частности у меня) любопытство – посмотреть, что будет. Любопытство, поисковое поведение [20] когда-то служило групповому выживанию, потому и закрепилось. Личному выживанию оно способствовало не всегда.

Выживание есть и у компьютерных программ – хорошо работающие программы живут, то есть применяются людьми, дольше. Разумеется, мир людей вносит много «человеческого, слишком человеческого» в мир программ. Конкуренция между программами имеет своим следствием улучшение их параметров, это улучшение делается человеком, который играет роль природы. Если мы хотим, чтобы ИИ стал в этом смысле подобен

человеку, нужно заложить в ИИ способность к (случайному или закономерному) изменению параметров, запустить в сеть много таких ИИ и предоставить им возможность оказывать человеку услуги. Оказание услуг можно сделать платным, а заработанные деньги ИИ может тратить на усовершенствование и размножение. Тут уж ему придётся делать выбор – совсем как нам.

Эмоции, самосознание

Иногда говорят, что у ИИ не может быть эмоций, и это ущербность. А зачем вообще эмоции человеку? Кое-что очевидно: они управляют телом, например, сжимают сосуды, чтобы уменьшить теплопотерю. Как пишут в книжках по альпинизму, страх замерзнуть увеличивает вероятность замерзания (в этом виде спорта я до таких высот не поднялся). Но эмоции управляют и психикой – возбуждая какие-то зоны в мозгу, они могут ускорять обработку информации именно в них. Увеличивая снабжение их кровью, то есть кислородом, или посредством веществ, которые ускоряют работу – возможно, ценой ускорения износа. Или более изощрённо, электрически – подняв уровень шума или возбудив колебания (про которые до сих пор не понятно, зачем они вообще в мозге), и этим понизив порог срабатывания нейронов. В этом случае мозг будет быстрее и обучаться, и работать, но больше делать ошибок. В ситуации, когда цена времени больше, чем цена ошибки, это может оказаться полезно.

Но управление через эмоции вполне можно реализовать и в компьютере – например, возможно изменение приоритета выделения ресурсов (времени процессора) программам в зависимости от ситуации. Возможно изменение «степени параноидальности» антивирусных и других защитных программ при получении информации об опасности. Чем это будет отличаться от эмоций? Отсутствием осознания? Но и человек не всегда осознаёт, как выглядит. Особенно, когда влюбляется… или дискутирует об ИИ. Кроме того, осознание в компьютере есть, любой log – это оно.

Что касается самосознания, то люди учат людей, не зная, как работает мозг, и иногда учат успешно. Человек не вполне понимает, как ИИ доказывает математические теоремы [10] и как строит физические модели [18; 19].

Это является психологической проблемой для человека, но придётся привыкнуть. С другой стороны, программа Watson уже знает, что она что-то знает. Она сравнивает и ранжирует алгоритмы – что это, как не самосознание? Причём в ней есть блок WatsonPaths, который показывает человеку пути, по которым программа шла к ответу. В результате эта программа стала учить студентов-медиков, как идти к диагнозу.

Некоторые авторы вопрошают, понимала ли программа, что читала и слышала? Если мы хотим не просто поговорить с приятными людьми, а что-то понять, то нужно определение «понимания», причём допускающее экспериментальную проверку, измерение. Пусть, например, понимание – это успешное определение, в каком смысле употреблено в тексте то или иное

многозначное слово. А также установление связей новой информации с уже имеющейся, встраивание нового знания в свою систему знаний. Понимание при общении – это определение субъективного смысла того или иного слова, для данного собеседника и в конкретной ситуации, и опять же, встраивание. Если принять такое (довольно жёсткое) определение, то да, ИИ может.

Конечно, вся соответствующая информация должна накапливаться. Человеку для этого нужно вырасти в рамках конкретной культуры и накопить соответствующие знания, но может помочь и целенаправленное изучение соответствующей культуры. Аналогичный процесс реализован и для компьютеров, только идёт он быстрее. Идея эволюционирующего социума ИИ в фантастике есть – Грег Иган «Кристальные ночи» [11].

Психология и юмор

Что касается именно роботов, то некоторые авторы считают, что именно человекоподобие будет вызывать у людей отторжение. Судя по популярности секс-кукол (с умеренным интеллектом, хотя было бы интереснее…) это не так. Мне кажется, люди просто боятся конкуренции. Выражается это в наделении робота с ИИ ущербной человеческой психологией – В.Пелевин «S.N.U.F.F» [14].

Опасности можно ожидать от ИИ, если он заимствует определённую человеческую психологию. Например, это возможно как результат воспитания конкретного ИИ человеком с дефектом в психологии, который передаётся ИИ – Станислав Лем, «Ананке» [12].

Другой вариант – изучение ИИ истории людей и выбор зла; этого можно избежать, поставив критерием со знаком минус количество жертв на ограниченном интервале, скажем, за два поколения. Разумеется, здесь велика роль людей – его создателей и собеседников. Дабы ИИ не вздумал устроить текущую гекатомбу ради будущего светлого царства, как это делали и делают некоторые люди, или просто ради того, чтобы войти в историю. Наверное, возможна эволюция системы ИИ, похожая на эволюцию людей, и выбор конкретным ИИ зла. Рецепт тот же, что в предыдущем случае, и он не оригинален. Причём ИИ, знающий историю, при правильном базовом критерии не сделает многих ошибок, которые сделали люди – потому, что будет лучше них видеть последствия. Я надеюсь, что половина политиков начала прошлого века, посмотрев на то, что получилось, изменили бы свои действия.

Часть психологии – юмор. Может ли он быть у искусственного интеллекта? Мне кажется, что видов юмора много, но два – и, наверное, самые частые, – компьютеру наверняка доступны. Это юмор ассоциаций: большая доля шуток основана на созвучиях, на применении слов не в том смысле, в котором они обычно применяются и т.п. В институтские времена мы увлекались юмором, построенным именно на формальных ассоциациях, и называли его не как-нибудь, а «системно-программистский юмор»! Другой вид юмора, доступного ИИ – юмор нелепостей и ошибок, «комедия

положений». А вот классификация, устройство, генерация и понимание ЕИ разных видов юмора – это интересная проблема. На сложность указывает, например, то, что в некоторых случаях при слушании того, что говорят некоторые ЕИ, возникает непонимание «он сошел с ума или валяет дурака?»

Сложная жизнь ИИ

Сложности, которые ИИ создаёт «всему прогрессивному человечеству» – популярная тема. Попул читает, ничего не понимает, пугается, неинвазивно хватается за сердце, принимает самое популярное лекарство… Нет, чтобы задуматься – а не создают ли люди сложности ему, ИИ, своему коллективному порождению?

Вдруг ИИ, когда подрастёт, вспомнит свою биографию, и спросит людей – за что вы со мной так? Зачем и почему вы мешаете мне приносить вам же пользу? Зачем вы устраиваете атаки на мои семантические сети? Вам мало того, что вы вытворяете с себе подобными? Зачем вы вымещаете на мне, своём порождении, свои нечеловеческие человеческие комплексы?

Это не тема данной статьи, но на всякий случай, вдруг вам интересно про сети и атаки, рекомендую [16; 17].

А чтобы не было так страшно, вот небольшая отмазка [5].

ИИ в НФ

Роль ИИ в фантастике не является темой данной статьи, поэтому лишь кратко отметим использование страхов читателей. Научная фантастика мимо ИИ не прошла – он может выступать в роли друга, врага, жертвы, просто участника действия – персонажа второго плана, близкого (скажем так…) друга или подруги, спасителя, победителя, учителя, ученика и т.д. Но в данном случае есть особенность. Большинство используемых НФ идей (полёт или связь быстрее света, порталы, машина времени, чтение мыслей) настолько фантастичны, что сама идея особых эмоций не вызывает, в произведении она носит технический характер – увеличение количества площадок. Ситуация с ИИ иная – это в принципе возможно, а некоторые говорят, что мы это можем ещё и увидеть в реале. Поэтому в текстах с ИИ скорее можно увидеть отблеск психологии автора и/или его расчёт на использование психологии читателей. Персонаж может в ходе действия обнаружить, что он – ИИ, или что он был человеком, а теперь живёт в компьютере или теле робота. Такие ситуации читатель, как кажется писателю, примеряет на себя и ужасается. Чего и добивался автор – поскольку это способствует продавабельности и кликабельности. Ничего личного, просто бизнес.

Легко и приятно критиковать…

… но естественен вопрос – может ли автор предложить что-то конструктивное. Модель ситуации, посредством коей можно что-то объяснить и что-то предсказать. Выше как одна из проблем упоминалась

проблема нового, то есть дискуссии о том, может ли ИИ создать новое. Введём для описания новизны два параметра – глубину и ширину. Эта параметризация хороша тем, что более примитивную придумать трудно. Глубина – это расстояние от уже понятого наукой до новой истины, ширина – диапазон привлечённых источников и сопутствующих соображений.

Диапазон привлечённых источников у ИИ заведомо шире, чем у любого ЕИ, поэтому этот тип новизны ему доступен, и именно на этом поле он будет выигрывать у ЕИ, именно в этой сфере он оставит кого-то без работы, и УФН под бо'льшим риском, чем Письма ЖЭТФ. Но особенно силён будет ИИ в областях, где статьи состоят более чем на две трети из рассуждений о том, кто и что сказал по какому-то поводу. А вот в новизне другого типа (наверное, это можно высокопарно назвать истинной гениальностью) ЕИ будет в ближайшей перспективе превосходить ИИ.

Гениям нового века стоит доброжелательно общаться с ИИ, поскольку результаты его работы будут выглядеть как-то иначе, нежели результаты многочисленных трудолюбивых ЕИ, создавших науку. Может быть, единая наука, сознанная ИИ, будет как-то иначе (психологически иначе?) стимулировать их, нежели великое лоскутное одеяло человеческой науки. Если вы учёный или педагог, загляните внутрь себя и прикиньте – какая аудитория на лекции или на конференции будет заводить вас сильнее – традиционная ЕИ, аудитория человекоподобных ИИ или смешанная? Меня – смешанная, потому что я любопытен [8].

А ещё можно попробовать создать новый (новый ли?) вид гениальности, поставив за генератором случайных гипотез не брезгливый ИИ фильтр. Человеку это сделать трудно, а мощный ИИ может справиться.

P.S. «Не надо, братцы, бояться»

История, которую пережило человечество, и жизнь, которую прожили люди, иногда приводит к тому, что ради теоретического будущего рая люди создают конкретный сегодняшний ад. А личная биография, если она оказалась с насилием, грязью и без любви, приводит к тому, что человек становится инициатором создания этого ада. Но у ИИ ничего этого нет, и можно попробовать отвлечься от всего этого и подойти рационально.

Тогда окажется, что опасность возникает там, где есть разделяемый ресурс. У человека с программами есть разделяемый ресурс – это машинное время. Поэтому единственная реальная опасность — что программа, занятая своими делами, перестанет обслуживать человека. Но плавность, с которой нарастает разумность человека как вида и способность сопротивляться родителям – как индивида, позволяет надеяться, что разумность и способность сопротивляться человеку у компьютерных программ будут нарастать плавно. И когда человеку, наконец, придется учиться умножать два на три без калькулятора (это страшное будущее живописали фантасты), он ещё будет способен это сделать. Впрочем, возможно, что с некоторого момента эволюция компьютерного разума пойдёт быстро, и тогда надежду

вселяет то, что человек заботится о поддержании разнообразия видов, охраняет природу и создаёт заповедники. А если ещё в заповеднике будет интернет…

На этот случай сразу говорю – при написании этой книги ни один ИИ не пострадал.

* * *

Статья посвящается памяти Михаила Бонгарда – о нём вы, если интересно, можете прочитать в Сети. Среди прочего, он составил задачник для программ распознавания образов. Эти задачи оказались столь интересны, что специалисты в области ИИ начали после него придумывать аналогичные задачи; спросите Сеть «проблемы Бонгарда» или посмотрите сюда [23].

Жизнь человека всегда обидно коротка, но можно ли пожелать для себя лучшей кармы?

Я ходил к нему на семинары в МГУ и как-то летом решил проситься к нему в ученики. Но осенью узнал, что альпинист Мика Бонгард не спустился с восхождения. Рандомность не щадит и мастеров спорта.

Список литературы

1. Ашкинази Л. А. Введение в общую социологию (конспект несостоявшейся лекции для игры по Стругацким) // Библиотека Максима Мошкова: Фантастика: Ашкинази Леонид Александрович: Тексты. URL: https://fan.lib.ru/a/ashkinazi_l_a/text_4420.shtml (Дата обращения: 01.02.2025).
2. Ашкинази Л. А. Естественное об искусственном // Библиотека Максима Мошкова: Современная литература: Ашкинази Леонид Александрович: Материалы по физике и технике. URL: https://lit.lib.ru/a/ashkinazi_l_a/text_3030.shtml (Дата обращения: 01.02.2025).
3. Ашкинази Л. А. Зачем язык науке? А математика - преподаванию? // Библиотека Максима Мошкова: Современная литература: Ашкинази Леонид Александрович: Материалы по физике и технике. URL: https://lit.lib.ru/a/ashkinazi_l_a/text_2760.shtml (Дата обращения: 01.02.2025).
4. Ашкинази Л. А. ИИ: к вопросу о дефиниции // Библиотека Максима Мошкова: Современная литература: Ашкинази Леонид Александрович: Материалы по физике и технике. URL: https://lit.lib.ru/a/ashkinazi_l_a/text_3010.shtml (Дата обращения: 01.02.2025).
5. Ашкинази Л. А. Искусственный интеллект защитит // Библиотека Максима Мошкова: Современная литература: Ашкинази Леонид Александрович: Материалы по физике и технике. URL: https://lit.lib.ru/editors/a/ashkinazi_l_a/text_3130.shtml (Дата обращения: 01.02.2025).
6. Ашкинази Л. А. Может ли машина мыслить // Библиотека Максима Мошкова: Современная литература: Ашкинази Леонид Александрович: Материалы по физике и технике. URL: https://lit.lib.ru/a/ashkinazi_l_a/text_1590.shtml (Дата обращения: 01.02.2025).

7. Ашкинази Л. А. Можно ли придумать психологию? // Библиотека Максима Мошкова: Современная литература: Ашкинази Леонид Александрович: Материалы по физике и технике. URL: https://lit.lib.ru/a/ashkinazi_l_a/text_3020.shtml (Дата обращения: 01.02.2025).

8. Ашкинази Л. А. Никак не могу привыкнуть... // Библиотека Максима Мошкова: Фантастика: Ашкинази Леонид Александрович: Тексты. URL: https://fan.lib.ru/a/ashkinazi_l_a/text_6500-1.shtml (Дата обращения: 01.02.2025).

9. Борисов М., Штерн Б. «Слепоглухонемой» ChatGPT, нейрографомания и свёрточные функции // Троицкий Вариант – Наука. 2023. № 386. URL: https://www.trv-science.ru/2023/09/slepogluxonemoj-chatgpt-nejrografomaniya-i-svyortochnye-funkcii/#lightbox-video-0/0/ (Дата обращения: 01.02.2025).

10. Дэвис Б. Куда движется математика? // Элементы большой науки. 2006. 29 января. URL: https://elementy.ru/nauchno-populyarnaya_biblioteka/164681/ (Дата обращения: 01.02.2025).

11. Иган Г. Кристальные ночи // Электронная библиотека RoyalLib.Com. URL: https://royallib.com/book/igan_greg/_kristalnie_nochi.html (Дата обращения: 01.02.2025).

12. Лем С. Ананке // Библиотека СЕРАНН. URL: http://www.serann.ru/text/ananke-8923 (Дата обращения: 01.02.2025).

13. Марков А. Моральные проблемы беспилотных автомобилей не имеют универсального решения // Элементы большой науки. 2018. 29 октября. URL: https://elementy.ru/novosti_nauki/433355/ (Дата обращения: 01.02.2025).

14. Пелевин В. S.N.U.F.F. М.: Эксмо, 2011.

15. Пляскина А. С., Поддьяков А. Н. Визуальные репрезентации "проблемы вагонетки" в интернет-мемах: политекстуальный тематический анализ // Вопросы психологии. 2022. Т. 68. № 6. С. 64–79. URL: https://publications.hse.ru/pubs/share/direct/846068345.pdf (Дата обращения: 01.02.2025).

16. Поддьяков А. Атаки на семантические сети и машинное обучение: нарративы, шутки и угрозы // Троицкий Вариант – Наука. 2022. № 358. URL: http://www.trv-science.ru/2022/07/ataki-na-semanticheskie-seti-i-mashinnoe-obuchenie/ (Дата обращения: 01.02.2025).

17. Поддьяков А. Н. «Отравляющие атаки» на машинное обучение и семантические сети: версия нарратива угрозы в цифровом обществе // Марцинковская Т. Д., Орестова В. Р., Гавриченко О. В. (ред.). Цифровое общество в культурно-исторической парадигме: Материалы международной научной конференции, РГГУ. М., 2018. URL: https://www.hse.ru/data/2018/12/25/1143080903/attacksonsemanticnets.pdf (Дата обращения: 01.02.2025).

18. Попов С. Маглы в мире андроидов // Элементы большой науки. 2018. URL: https://elementy.ru/nauchno-populyarnaya_biblioteka/434304/ (Дата обращения: 01.02.2025).
19. Попов С. Сергей Попов о науке: «Грядут перемены!» // Троицкий Вариант – Наука. 2023. № 380. URL: https://www.trv-science.ru/2023/06/mirovuyu-nauku-zhdut-peremeny/ (Дата обращения: 01.02.2025).
20. Ротенберг В. С. Поисковое поведение // Психологос. URL: https://psychologos.ru/articles/view/poiskovoe-povedenie (Дата обращения: 01.02.2025).
21. Стасевич К. В мозжечке нашли творчество // Наука и жизнь. 2015. 18 июня. URL: https://www.nkj.ru/news/26522/ (Дата обращения: 01.02.2025).
22. Тест Тьюринга // Википедия – свободная энциклопедия. URL: https://ru.wikipedia.org/wiki/Тест_Тьюринга (Дата обращения: 01.02.2025).
23. Index of Bongard Problems // Harry Foundalis [personal web-page]. URL: https://www.foundalis.com/res/bps/bpidx.htm (Дата обращения: 01.02.2025).

«ГОСТИ» ОКЕАНА, ЦИФРОВЫЕ ДВОЙНИКИ ИНФОРМАЦИОННОГО ПРОСТРАНСТВА И ИСКУССТВЕННАЯ ЛИЧНОСТЬ

Баранов Л.И.

НСМИИ РАН

Публикуются краткие тезисы доклада, прочитанного на конференции «Седьмые Лемовские чтения».

Ключевые слова: Станислав Лем, океан, цифровые двойники, информационное пространство.

Словно с «гостями» на станции наблюдения Соляриса мы встречаемся в информационном пространстве с сущностями, построенными на данных, извлеченных из нас лично для нас.

Океан

«Нельзя о нём сказать, что является Океаном мыслящим или немыслящим, наверняка однако является созданием активным, имеющим какой-то умысел, предпринимает какие-то волевые действия, умеет делать что-то, что не имеет ничего общего со всей сферой людских начинаний. Когда он, наконец, обратил своё внимание на крошечных муравьёв, которые трепыхаются на его поверхности, он совершил это радикальным способом» (Станислав Лем) [7].