

Научно-исследовательская работа

Философия

**Искусственный интеллект как этическая проблема
развития науки и техники**

Выполнил: Барахтин Николай Николаевич

Учащийся 221гр.- отделения ЭТ,

*Санкт-Петербургского техникума железнодорожного транспорта, Россия,
г. Санкт-Петербург*

Колупаева Александра Юрьевна

научный руководитель,

*Санкт-Петербургский техникум железнодорожного транспорта, Россия, г.
Санкт-Петербург*

Введение

Искусственный интеллект (ИИ; англ. *artificial intelligence, AI*) — свойство интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека.

Это определение искусственного интеллекта, дал Джоном Маккарти в 1956 году на конференции в Дартмутском университете. Следует отметить, что оно не связано напрямую с пониманием интеллекта у человека. Согласно Маккарти, исследователи ИИ 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».

Простыми словами, под ИИ подразумевается машина, способная выполнять те же задачи что и мы, на всех уровнях жизни.

Немного об истории:

История искусственного интеллекта как нового научного направления начинается в середине XX века. К этому времени уже было сформировано множество предпосылок его зарождения: среди философов давно шли споры о природе человека и процессе познания мира, нейрофизиологи и психологи разработали ряд теорий относительно работы человеческого мозга и мышления, экономисты и математики задавались вопросами оптимальных расчётов и представления знаний о мире в формализованном виде; наконец, зародился фундамент математической теории вычислений — теории алгоритмов — и были созданы первые компьютеры.

Возможности новых машин в плане скорости вычислений оказались больше человеческих, поэтому в учёном сообществе зародился вопрос: каковы границы возможностей компьютеров и достигнут ли машины уровня развития человека? В 1950 году, английский учёный Алан Тьюринг, пишет статью под названием «Может ли машина мыслить?», в которой описывает процедуру, с помощью которой можно будет определить момент, когда машина сравняется в плане разумности с человеком, получившую название теста Тьюринга.

Часть 1. Этические проблемы

Какие же этические проблемы могут возникнуть при интеграции ИИ в жизнь человека. И самое главное - как их решить. Рассмотрим на примерах, которые показались мне более важными и значительными.

-Безработица

-Безопасность и контроль

-Права ИИ

-Гуманность Идентичности

1) Безработица. Конечно же, с появлением интеллекта способного выполнять простые задачи, запустится глобальная автоматизация, что приведёт к тотальному вымиранию профессий, связанных с физическим трудом, которые были так необходимы в индустриальную эпоху. Даже сейчас, в мире около 2,9 млрд. людей заняты в промышленности и сельском хозяйстве. Это около 40% населения Земли. Этот процесс естественно повлечёт за собой появления новых профессий, где будет использоваться когнитивный труд (сфера услуг). Но что станет с теми людьми, которые не смогут переключиться и найти себе работу? Естественно, вспыхнут забастовки, хаос и анархия. Как же этого избежать? Я считаю, что эта проблема уходит корнями во всем нам знакомое с детства неравенство, а именно: неравенство доступа к ресурсам, в первую очередь к интеллектуальным ресурсам. Ведь если бы все люди имели возможность развиваться одинаково, то к зрелому возрасту высокую квалификацию получили бы те люди, которые этого правда хотели. Те люди, которые старались и развивались. А высококвалифицированных работников не заменяют роботами, роботами заменяют низко квалифицированных рабочих, и они сами в этом виноваты, в этом виновата их собственная лень и безответственность. Но так как наш мир не идеален с самого начала, то тут уж ничего не поделаешь. Так вот, что бы избежать этой проблемы нужно либо постепенно внедрять систему ИИ, что пока не дало результата, либо кардинально менять систему образования и распределения, к чему я всех вас призываю.

2) Переходим к следующей проблеме. Если же получится без тяжких последствий интегрировать ИИ в человеческую систему, то, как обезопасить человека от машины или в некоторых случаях машину от человека? Когда я пытался разобраться в этом вопросе, то обнаружил так называемые «Три закона робототехники», которые были опубликованы в рассказе «Хоровод» и **сформулированы Айзеком Азимовым, американским писателем-фантастом** популяризирующим науку. Согласно рассказу, эти три закона являются некими обязательными правилами поведения для роботов. Вот их содержание:

1. Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред.
2. Робот должен повиноваться всем приказам, которые даёт человек, кроме тех случаев, когда эти приказы противоречат Первому Закону.
3. Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому или Второму Законам.

Этакая идеальная полиция получается. Но естественно, эти три установки должны находить в надёжнейшей защите. Потому что найдутся те, кто захочет их переписать, это не сложно, учитывая сегодняшний уровень киберприступности (другой вопрос - как сам взломщик собирается контролировать робота?). Но представим, что есть такая защита, при которой эти вопросы нельзя видоизменить или удалить. Можно ли при таком раскладе обойти систему? Сам Айзек Азимов в 1986 году в романе «Роботы и Империя» ответил на этот вопрос, предложив Нулевой закон робототехники, его формулировка почти не отличается от первого:

0. Робот не может причинить вред человечеству или своим бездействием допустить, чтобы человечеству был причинён вред.

В чём же тут лазейка, спросите вы? В масштабности. Человечество убивает само себя уже очень долго. Люди загрязняют атмосферу, из-за этих

загрязнений происходят катаклизмы, а также войны, насилие. И роботу в этой ситуации, что бы защитить человечество нужно одновременно устранить ту его часть, которая это самое человечество разрушает. Сложность для робота заключается в выборе. Ведь у каждого человека свои нравы, он сам для себя определяет что хорошо, а что плохо. А как роботу для себя определить что правильно, а что нет. Так вот, какое, в этой ситуации, принять решение, не знаю даже я. Тем самым прихожу к выводу, что, с точки зрения морали, прийти к консенсусу в вопросе безопасности нельзя.

3) Если же удастся создать ИИ, внедрить его в нашу систему и добиться более- менее идеальной безопасности, то, что тогда делать с нашими межличностными отношениями. Нужны ли роботам такие же права, как и нам? Почему нет? У нас примерно равный интеллект, это нормально иметь одни и те же юридические привилегии. Но их не будет, хотя бы сразу. «Почему?». А потому, что люди склонны к дискриминации, причём абсолютно разного рода. Люди не любят, когда кто-то отличается от их этнической группы, люди всегда начинают бояться чего- то нового и не похожего на то, что они видели прежде; страх перерастает в злость, а злость в агрессию и жестокость. В истории много таких примеров: рабство на почве расизма существовало в США на протяжении 249 лет. Рабство на почве сословного неравенства существовало в Российской Империи на протяжении 212 лет. Дискриминация Женщин продолжалась 181 год (с начала 19 века «Индустриализация» по 1981 г. Конвенция о ликвидации всех форм дискриминации в отношении женщин). С роботами будет та же ситуация, их будут унижать, оскорблять просто по факту того, что они состоят не из мяса. Их права не будут признаваться из-за того, что люди будут считать их недостаточно человеческими. Они будут видеть в них машины, а не сущность, обладающую собственной особенностью черт и устоев, они будут видеть в них инструменты, просто потому, что у нас разный каркас. Естественно, что бы признать робота равным человеку, в нём должны быть способность испытывать боль и сопереживание, самосознание и возможность смотреть на

вещи с точки зрения морали и принимать соответствующие решения. Такой ИИ и должен получать права (хотя не все люди способны на сострадание и бескорыстие, честность и порядочность... а права получают все).

Все люди – личности, но не все личности – люди. ©Линда Макдональд-Шленн, специалист по биоэтике Калифорнийского университета Монтерей-Бэй.

4) Гуманность Идентичности. Что это и о чём - мы будем говорить. Если коротко, то это Трансгуманизм. Тут сразу нужно понять, что Культурными предпосылками трансгуманизма оказываются отказ от субстанциальной парадигмы... Я не буду изрекаться такими выражениями, выразаться такими изречениями, потому что это непонятно, скучно и это вообще не выговоришь... Но речь всё же пойдёт о довольно непрстом понятии, поэтому давайте вникать. Трансгуманизм - это философская концепция, а также международное движение, поддерживающее использование достижений науки и технологии для улучшения умственных и физических возможностей человека. То есть - это устранение таких побочных эффектов нашей жизни как страдание, болезни, старение и смерть. В этом состоят их цели и задачи, также как и поддержание технического прогресса, расширение свободы каждого отдельно взятого человека путём внедрения механизмов в его организм, а также борьба против антипрогрессистов и антимодернистов. Вы представьте насколько это круто - отдаление, а в идеале — отмена старения и смерти человека, мы даём ему право самому решать, когда умирать и умирать ли вообще, радикально уменьшаем страдания и увеличиваем уровень счастья. И это всё ждёт нас в ближайшем будущем, потому что сейчас развиваются такие дисциплины как Нейропротезирование, Выращивание органов, Клонирование, Искусственные органы, Психопрактика, экзоскелет. Это даст огромный толчок к формированию таких наук как Генная инженерия, Загрузка сознания, молекулярная нанотехнология, Крионика. **Это новая ступень эволюции.**

Я не нашёл адекватной критики этому превосходному движению. Единственное, что доступность таких методов опять будет очень узкой среди населения. И мы пришли к тому, с чего начали - НЕРАВЕНСТВО. Его порождает, человеческий эгоизм, невежество, упрямство, нарциссизм, самолюбие, лицемерие, алчность, властолюбие, безразличие, наглость. А ещё ограниченность, расточительность, мстительность, мелочность, тупость, грубость, трусость, лживость, зависть. ВОТ главный корень всех проблем!

Можно ли их решить?

Хороший вопрос, но сейчас я хотел бы немного отвлечься от темы. В конце мы к этому вернёмся. Я бы хотел поговорить ещё вот о чём. Мы уже знаем, с чем имеем дело, и что за этим всем стоит. Но мы не разобрались в том, как же создать этот самый Интеллект?

Часть 2. Создание

Для начала, нужно иметь представление о том, что мы хотим создать. И получается, что мы пытаемся воспроизвести **сознание**. Так вот вопрос: где же находится человеческое сознание? Головной мозг. Хорошо, а где в мозгу сознание? Головной мозг – это совокупность нервных клеток. Если брать отдельно клетку (нейрон),

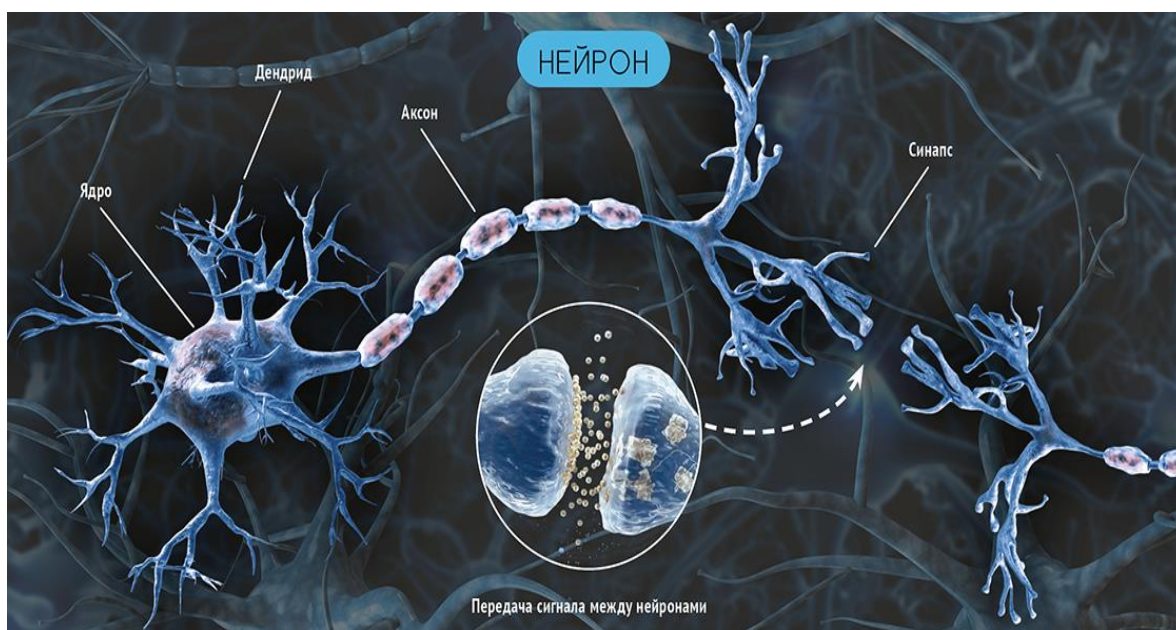


Рисунок 1. Строение нейрона

В середине нейрона ядро, вокруг тело клетки, от него исходят дендриты, они переходят в аксон, который соединяет эту клетку со следующей. Так где в этой структуре нейрона находится сознание? В ядре? В митохондриях? Конечно, это неправильно так ставить вопрос. Ведь сознанием можно считать всю сеть из множества нейронов. Согласитесь, нельзя ответить на вопрос «Что такое интернет?», рассматривая отдельный компьютер. Кстати, это хорошая аналогия с мозгом - интернет. Сеть компьютеров - это и есть интернет. А сеть нейронов – это и есть сознание. Но... на вопрос «что такое Интернет?» есть ответ и сложность в том, что он не один. Например: Интернет – это структура, образованная компьютерами по всему миру, соединенными друг с другом с помощью общего протокола. Также можно сказать, что любой отдельный компьютер – часть Интернета. А еще можно сказать, что люди, сидящие за компьютерами, – тоже часть Интернета. Опять же всё раскладывается на составные части и всё зависит от того, с какой стороны мы смотрим. Так как же всё-таки связано сознание человека и всемирная сеть? Если связи между серверами, роутерами, компьютерами и мобильниками образуют сеть, и эта сеть – «форма» Интернета. Тогда что такое его суть? Так вот, суть это то, что течёт по этим сигналам, **электричество** - уникальная структура материи, универсальная среда как для нас, так и для него. Суть Интернета – в том, как именно они распространяются по структуре и передают информацию... Давайте пока объясним так.

Так вот, сеть из сотен миллиардов соединенных вместе мозговых клеток... это форма сознания, тогда суть сознания есть информация, которая передается вдоль клеток в виде потока электрических импульсов. Можно пойти дальше, разобрав, что же такое импульс? Для этого нужно повнимательнее посмотреть на синапс, место соединения нейрон-нейрон, единственный триггер, работающий на веществах - нейромедиаторах. Он

необходим для передачи сигналов между нервными клетками; можем ли мы сказать, что это и есть суть сознания? От этих размышлений быстро устаешь. Перерыв множество статей, это последнее, к чему я пришёл. На самом деле наука не знает что такое сознание, и дальше этого мы пока не продвинулись. Это можно сказать тупик проверенных знаний. И когда заканчивается сфера проверенных знаний, открывается область собственных. И вот я решил создать собственную теорию, основанную на «Квантовой динамике мозга». Кажется, ее сформулировал какой-то английский ученый в конце прошлого века. Покопавшись в ней, я мало чего понял.

Но вот результат моих рассуждений. Сейчас должен предупредить, следующее, что вы услышите - это всего лишь мои догадки. Мы только что говорили об устройстве клетки мозга, так(?). У этой клетки есть каркас, который поддерживает ее структуру. Кажется, он называется «микротрубочками». Они не только поддерживают клетку, они образуют нечто вроде черепа. Как бы Мозг внутри клетки мозга. Этот «черепа» имеет форму трубки – пустотелого цилиндра. Ну, конечно, очень маленького... его толщина измеряется нанометрами, но он не пустой. Там, в этой трубке, кое-что есть. И скорее всего это – Свет – Частица света... ее называют «эфемерный фотон». Фотон – это квант. Он существует в состоянии неопределенности, постоянно флуктуирует, пульсирует в соответствии с теорией вероятности. Эти пульсации...на мой взгляд и есть суть человеческого сознания. Я бы назвал эту частицу FluctLight (квантовая сущность мозга), потому что с английского Fluctuating Light - **Колеблющийся свет**. Думаю, даже можно считать это душой человека. Это теория довольно сильно упрощает основу работы человеческого мозга. Получается, мы берём мозг не как орган с отдельными частями, а единую «сеть души». В принципе, в будущем, можно построить машину (STL «Soul translator»), которая будет воздействовать и взаимодействовать напрямую с «душой», некая оцифровка мозга получается.

Вот только с реализацией будет туго. Ведь что бы воссоздать такую сложную структуру, даже с передовыми технологиями, уйдёт лет 20. Ведь нужно действовать с особым подходом. Различают два подхода при создании ИИ: сверху- вниз и снизу-вверх. Сверху- вниз, суть в том, что бы заложить знания и опыт в программу и путём обучения приблизить интеллект к настоящей разумности. Все аналоги ИИ в мире работают именно так, однако они могут адекватно реагировать только на знакомые ситуации и полностью разумным его назвать нельзя. Второй способ - это построение снизу- вверх, это - имитация мозга. В начале, создаётся искусственная модель сети из триллиона связанных между собой клеток, а затем её пытаются научить думать. В принципе, если создать носитель по вместимости сравнимый с мозгом, некую квантовую матрицу, то можно просто скопировать или как я уже говорил оцифровать чей-то интеллект. Тогда, что-то подобное уже должно было существовать. И тут я понял свою главную ошибку, она заключается в том, что между копией и ИИ есть огромная непреодолимая разница. Копия не может выдержать самого факта своего существования, признать, что сама она является копией чьих- то воспоминаний и образа мышления просто непостижимо, и в результате она просто убьёт себя. Есть другой способ - менее легальный. Скопировать и выращивать души новорождённых, правда, мы сталкиваемся с ещё одной проблемой, как создать среду (полноценный симулятор современного мира) для выращивания. Решения есть VRMMORPG с открытым миром. За играми будущее. Таким образом, мы можем создать ИИ, только в этом случае у нас с ними будет одно отличие, у них не будет воли, потому что у них не будет инстинктов к выживанию, той части нашего мышления, в которой ещё не существует понятия морали. И если ввести кодекс запретов, программа будет исполнять их беспрекословно, на то она и программа. Не будет воровства, убийств, неравенства в отдельно взятом, небольшом мире. Одним словом, утопия. И получается, в пределах сегодняшних возможностей или возможностей ближайшего 30-летия, такой мир возможно будет создать, в

качестве программы,... виртуальной игры. А в реальности нет. В первую очередь из-за ограниченности ресурсов и первобытных инстинктов, которые у нас сформировались. А самое главное и обидное, что все знают как изменить мир, все знают, что нужно начать сотрудничать и избавиться от предрассудков. Но никто не знает, почему он не меняется. А не меняется он потому, что люди не меняются, не идут на уступки, не производят усилия над собой - они ждут, пока весь этот мир изменится, а не они.

Заключение:

Поэтому, чтобы решить проблемы, связанные с ИИ, нужно с начала решить проблемы связанные с человеком, а это, увы, невозможно.

Список литературы:

1. Интернет ресурс Википедия:

1.1 Статья о ИИ

1.2 Статья о «Трёх законах робототехники»

1.3 Статистические статьи.

2. Ранобэ «Мастера Меча Онлайн: Алисизация»

2.1 Сайт animeGo.org