

СВОБОДА ВОЛИ: РЕАЛЬНОСТЬ, ВОЗМОЖНОСТЬ ИЛИ ИЛЛЮЗИЯ?

Панев А.С.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Ратникова И.М. – канд. филос. наук

Статья посвящена обоснованию тезиса об отсутствии свободы воли у Искусственного Интеллекта и человека.

В наши дни Искусственный Интеллект (далее – ИИ) повсюду: алгоритмы YouTube, голосовые помощники, автопилот для автомобилей. С каждым днем сфер применения ИИ становится все больше, но зачастую мы об этом даже не догадываемся. К примеру, цены на недвижимость высчитываются ИИ, а не человеком.

В массовой культуре широко тиражируется позиция, согласно которой допускается возможность появления свободы воли у ИИ. В этом случае он сможет захватить мир и поработить человечество. Данное утверждение представляется весьма сомнительным. ИИ может разрушить мир только в двух случаях: а) если поставлена прямая цель захватить или разрушить мир; б) если поставлена какая-то цель, которая может быть наиболее эффективно достигнута путём разрушения мира.

Последний вариант можно пояснить следующим примером. Программист ставит андроиду цель научиться поливать оранжерею так, чтобы влажность почвы всегда была в допустимых пределах. При этом схема полива должна потреблять минимальное количество электричества. Пусть за показания влажности почвы отвечают датчики в земле. ИИ может догадаться положить датчики в закрытую ёмкость с почвой необходимой влажности. В этом случае, несмотря на выполнение основной цели (влажность в допустимых пределах и малый расход электроэнергии), все растения погибнут. Таким образом, ИИ не сможет нанести вред человеку, если мы будем правильно формулировать цели.

Более вероятный сценарий развития событий начинается уже сейчас. ИИ всё больше и больше облегчает жизнь человеку в самых различных сферах. Так, в будущем появятся андроиды, которых нельзя будет внешне и по интеллекту отличить от человека. Для того чтобы избежать путаницы и различного рода недоразумений, необходимо будет присвоить андроидам отличительный признак (к примеру, горящую лампочку во лбу). При этом надо понимать, что это сходство будет лишь внешним и появление свободы воли у андроидов невозможно.

В то же время для многих людей, не понимающих принцип работы ИИ, использование андроидов в качестве рабочих, возможно, будет выглядеть как рабство. Данное обстоятельство может выступить в качестве основания массовых протестов. Например, уже сейчас людям жалко роботов. В социальных сетях все чаще и чаще можно встретить сообщения подобного плана: "Подарил маме робот-пылесос, но она его практически не использует, потому что ей его жалко". И это при том, что современные роботы-пылесосы пока даже внешне не похожи на живых существ. Зачастую люди склонны сочувствовать неживому объекту, приписывая ему качества живого существа: сознание и способность чувствовать боль и эмоции.

Представьте, что будет, если такие роботы начнут разговаривать и вести себя так же, как и люди. Должны ли мы их жалеть, должны ли мы верить в то, что все сказанные ими слова идут не из алгоритмов нейронных сетей, а из некой души, заточённой в микросхеме? Поверхностный ответ на этот вопрос: то, что ведёт себя так, будто у него есть сознание, обладает им. В то же время даже если андроиды и будут неотличимы от человека ни по внешности, ни по поведению, это всё ещё будет программа в искусственной плоти. Главной целью этой программы будет являться создание иллюзии того, что у андроида есть душа. Если же мы поверим в эту иллюзию, значит, алгоритм справился со своей задачей.

Разберем основные принципы работы ИИ. Есть входные сигналы, внутренние обрабатывающие "нейроны", их чаще всего много слоёв, и выходные сигналы (Рисунок 1). Сигнал передаётся слева направо, проходя через нейроны, каждый из которых получает сигналы и выдаёт результат согласно простейшим математическим формулам. Здесь всё подчинено принципам детерминизма, здесь нет места свободе воли или случайностям. Другими словами, при одинаковых входных данных одна и та же нейронная сеть выдаст всегда один и тот же результат. Человеческий мозг устроен таким же образом.

Ответим на вопрос: есть ли что-либо общее у месячных, опухоли мозга, злоупотребления сладостями и анаболическими стероидами? На первый взгляд, нет. В то же время именно эти факторы успешно использовались в суде для объяснения поведения убийц. Например, в случае со сладостями мужчине неимоверно сильно сократили срок за двойное убийство. Понятие "свобода воли" включает в себя ответственность за совершаемые действия. Получается, суд согласился с тем, что при определенных условиях мы можем частично или полностью терять свободу воли, а наши

поступки являются лишь результатом некоторых биохимических реакций, имевших место в нашем организме. Но кто сказал, что когда состояние мозга, уровень гормонов и сахара в крови в норме, то мы в ответе за свои поступки, а когда уровень сахара превышает допустимые значения, мы резко теряем управление над нашими действиями? Эта утрата контроля не может происходить резко: на самом деле, как представляется, ее вообще не происходит, поскольку свободы воли у нас не было уже изначально. К примеру, возьмём какое-либо простое действие. Пусть человек пытается согнуть палец. Мозг посылает сигнал, командуя мышцам сокращаться. Абсолютно незначительное, в целом безобидное действие, какие тут вообще могут быть последствия? Весьма серьезные, если в руках у человека пистолет, то есть действие всё-таки не такое уж безобидное.

Нейробиология позволяет найти конкретный нейрон в моторной коре, который отправил сигнал и заставил палец согнуться (Рисунок 2). Можно также найти нейроны в так называемой премоторной коре, которые послали сигнал в моторную, откуда нейрон послал сигнал в руку. Эти нейроны, в свою очередь, получили сигнал от фронтальной коры, а она от префронтальной, а та в свою очередь от участков мозга, которые отвечают за эмоции [1].

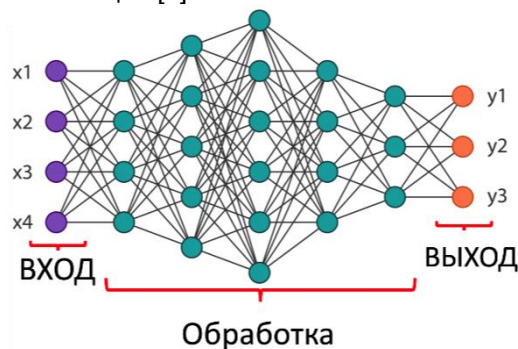


Рисунок 1 – Нейронная сеть

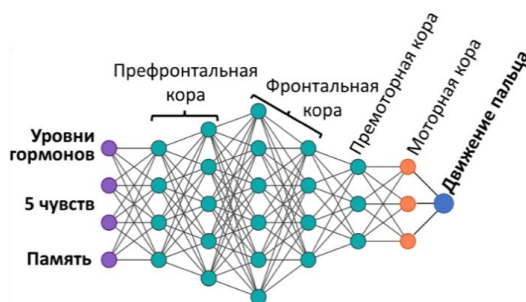


Рисунок 2 – Нейроны в мозге человека.

Зададимся весьма правомерным вопросом: существует ли нейрон, который стоит у истоков, который сработал бы сам, есть ли точка, где потенциал действия возник из ниоткуда, без внешних воздействий, или нейрон, который вдруг решил нарушить все известные физические законы, внезапно посылая сигнал без предшествующей этому причины? Пусть нам покажут нейрон, который способен на такое, и у нас появилась бы нейробиологическая основа свободы воли. Но пока этого не случилось. Не бывает причины без причины. Ничего в биологии нашего поведения не происходит просто вдруг. За каждым событием обнаруживается целая история [2].

Любой человек, какие бы у него ни были принципы, сидя за рулём машины, которая вот-вот неизбежно врежется в столб, направит машину левее, чтобы получить меньшие повреждения, чем человек, сидящий справа. Как бы нам ни казалось, что мы что-то решаем, мозг любого из нас при одинаковых параметрах и строении выдаст одно и то же действие. Случись у кого-либо такая же опухоль в отделе мозга, отвечающем за агрессию, как у убийцы, который ещё неделю назад был спокойным семьянином, и он бы тоже совершил преступление.

К настоящему времени было проведено большое количество невероятно важных исследований. Обратимся к одному из них. Так, ученые рассмотрели более 5000 судебных решений, принятых непосредственно на заседании, на предмет того, от чего зависит, отпустят ли заключенного досрочно или нет. Оказалось, что единственным показателем, который способствовал такому прогнозу – это то, сколько часов прошло с тех пор, как судья поел в последний раз. Если дело рассматривали сразу после обеденного перерыва, то шансы на досрочное освобождение составляли приблизительно 60 процентов, если же через два часа, то они стремились к нулю. Данное исследование подтверждено большим количеством фактического материала. Допускаю, что если попросить судью обосновать принятое решение о возможности досрочного освобождения, он вряд ли скажет про уровень сахара в крови, но приведет аргументы из области юридической литературы и практики.

57-я Научная Конференция Аспирантов, Магистрантов и Студентов БГУИР, Минск, 2021

Таким образом, мы принимаем огромное число решений без своего собственного ведома. Современная нейробиология убедительно доказывает, что когда мы совершаем обдуманый, как нам представляется на первый взгляд, выбор, обеспечиваемый наличием у нас свободы воли, оказывается, что на самом деле все уже было решено на несколько секунд раньше, а сознательная, рациональная часть мозга затем просто находит этому разумное объяснение *post factum*. Свободой воли, другими словами, не может обладать ни человек, ни ИИ.

Список использованных источников:

1. Курс "Биология поведения человека" [Электронный ресурс] / Стэнфордский университет. – Режим доступа: <https://youtube.com/playlist?list=PL8YZyта552VcePhq86dEkohvoTpWPuauk> – Дата доступа: 05.06.2020.
2. Интервью с Робертом Сапольски [Электронный ресурс] / Р.Сапольски. – Режим доступа: <https://youtube.com/playlist?list=PL8YZyта552VcfR6U49lyqClcDfpO65OPp> – Дата доступа: 12.10.2020..