

ДИСКУССИИ / DISCUSSIONS

Редактор рубрики П. А. Кабанов / Rubric editor P. A. Kabanov

Научная статья

<https://doi.org/10.21202/2782-2923.2024.1.264-279>

УДК 004.8:159.9:343:347.9

М. Ди Сальво¹

¹ CrossMediaLabs, г. Неаполь, Италия

Искусственный интеллект и киберутопизм в правосудии. Почему ИИ – не интеллект; борьба человека за собственное выживание

Переводчик Е. Н. Беляева

Микель Ди Сальво, доктор права, Неаполитанский университет имени Федерико II; CrossMediaLabs
E-mail: Mik.disalvo@gmail.com
ORCID: <https://orcid.org/0000-0002-9531-0591>

Аннотация

Цель: рассмотрение онтологических различий между человеческим и искусственным интеллектом и структурных расхождений на уровне дефиниций.

Методы: диалектический подход к познанию социальных явлений, позволяющий анализировать их в историческом развитии и функционировании в контексте совокупности объективных и субъективных факторов, что предопределило следующие методы исследования: формально-логический и социологический.

Результаты: феномен ИИ подвергнут всестороннему анализу в контексте киберутопизма и киберреализма, начиная от основателей теории искусственного интеллекта (Тьюринга, Минского, Бернштейна, фон Неймана) и заканчивая недавними работами Макса Тегмарка. Показано, что открытия и гипотезы нейробиологии не позволяют определить ИИ в качестве интеллекта по человеческим критериям. В приложение включены три короткие заметки, подводящие итог обсуждению: 1. О сознании машин. 2. О теории утопической кибернетической занятости и системы вознаграждения. 3. Обсуждение израильского исследования «Голодный судья судит строже».

Научная новизна: на основе теорий множественного интеллекта (Гарднер) и социального интеллекта (Торндайк, Гоулман) автор предлагает определение интеллекта как более сложной системы, чем та, которая может быть воспроизведена искусственными нейронными сетями, особенно в отношении взаимодействия животного и окружающей среды. Короткие заметки в приложении показывают неопределенность и риски, которые могут возникнуть в связи с широким использованием искусственного интеллекта в роли судей.

Практическая значимость: через корректное определение человеческого интеллекта автор приходит к определению искусственного интеллекта. За мифом об искусственном интеллекте обнаруживаются его пределы и объективные ограничения, которые необходимо предусмотреть, чтобы спасти самое ценное, что есть у человечества, – его человеческую сущность.

© Di Salvo M., 2024

Ключевые слова:

искусственный интеллект, правосудие, уголовное право, нейробиология, теория вычислений

Статья находится в открытом доступе в соответствии с Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), предусматривающем некоммерческое использование, распространение и воспроизводство на любом носителе при условии упоминания оригинала статьи.

Как цитировать статью: Ди Сальво, М. (2024). Искусственный интеллект и киберутопизм в правосудии. Почему ИИ – не интеллект; борьба человека за собственное выживание. *Russian Journal of Economics and Law*, 18(1), 264–279. <https://doi.org/10.21202/2782-2923.2024.1.264-279>

Scientific article

M. Di Salvo¹

¹ *CrossMediaLabs, Naples, Italy*

Artificial Intelligence and the cyber utopianism of justice. Why AI is not intelligence and man's struggle to survive himself

Translator *E. N. Belyaeva*

Michele Di Salvo, Doctor in Law, Università Federico II di Napoli;
CrossMediaLabs
E-mail: Mik.disalvo@gmail.com
ORCID: <https://orcid.org/0000-0002-9531-0591>

Abstract

Objective: to show the ontological differences between human and artificial intelligence and address structural divergences at the definitional level.

Methods: dialectical approach to cognition of social phenomena, allowing to analyze them in historical development and functioning in the context of the totality of objective and subjective factors, which predetermined the following research methods: formal-logical and sociological.

Results: a cross-cutting analysis was applied to the phenomenon of AI between cyber utopianism and cyber realism. Starting from a quote by Max Tegmark, the theory of artificial intelligence is reconstructed by the theorists who founded the discipline (Turing, Minsky, Bernstein, von Neumann) and it is discussed why – in light of the discoveries and assumptions of neuroscience – it is not possible to define it as intelligence according to human criteria. Three short notes are included in the appendix that complete the discussion: 1. on the consciousness of machines 2. on the theory of utopian cyber employment and remuneration 3. “The hungry judge is more cruel” (discussion on an Israeli study).

Scientific novelty: through the examination of multiple types of intelligence (Gardner) and social intelligence (Thorndike, Goleman), a more complex definition of intelligence is proposed than that which can be replicated by artificial neural networks, especially in relation to the interaction between animal and environment. Three short messages highlight the uncertainty and risks that may arise from the rampant use of artificial intelligence as judges.

Practical significance: starting from a correct definition of human intelligence, the author comes to the definition of artificial intelligence. Beyond the myth of AI, we discover its limits and the objective limitations we must provide for in order to save the most precious asset we have: mankind.

Keywords:

artificial intelligence, justice, criminal law, neuroscience, computational theory

The article is in Open Access in compliance with Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), stipulating non-commercial use, distribution and reproduction on any media, on condition of mentioning the article original.

For citation: Di Salvo, M. (2024). Artificial Intelligence and the cyber utopianism of justice. Why AI is not intelligence and man's struggle to survive himself. *Russian Journal of Economics and Law*, 18(1), 264–279. <https://doi.org/10.21202/2782-2923.2024.1.264-279>

Введение

Мах Tegmark в своей концептуальной книге *“Life 3.0 – Being human in the era of artificial intelligence”* («Жизнь 3.0 – как быть человеком в эпоху искусственного интеллекта», 2018) предложил сценарий применения ИИ к области права. Он пишет:

«Какие первые ассоциации приходят вам в голову, когда вы думаете о судебной системе в вашей стране? Если это длительная волокита, высокие расходы и иногда несправедливость, то вы не одиноки. Разве не было бы замечательно, если бы в первую очередь вы подумали об эффективности и справедливости?» (Tegmark, 2018, p. 136).

Помня об этом тезисе, углубимся в лежащий в его основе метод:

«Поскольку юридический процесс можно абстрактно рассматривать как вычислительную задачу, в которой вводится информация о доказательствах и законах и выдается решение, некоторые ученые мечтают полностью автоматизировать его с помощью робосудей. Это системы искусственного интеллекта, которые будут неустанно применять высокие юридические стандарты в каждом решении, не поддаваясь человеческим ошибкам, таким как предвзятость, усталость или отсутствие новейших знаний» (Tegmark, 2018, p. 136).

В главе «Логика и мышление» в книге Steven Pinker *“Rationality: What it is, why it seems scarce, why it matters”* («Рациональность: что это такое, почему она представляется недостаточной и почему она имеет значение») (Pinker, 2022) всесторонне разобраны такие понятия, как формальные и неформальные заблуждения.

Имеющиеся предпосылки приводят к негативным последствиям:

«Впервые в истории робосудьи в принципе могли бы обеспечить всеобщее равенство перед законом. <...> Их можно запрограммировать так, чтобы все они были одинаковыми и относились ко всем одинаково, применяя закон на основе прозрачности и настоящей объективности... Робосудьи могли бы также устранить те человеческие предубеждения, которые являются случайными, а не преднамеренными. Еще один недостаток человеческих судей заключается в том, что им не хватает времени на изучение всех деталей дела. В отличие от них робосудей можно легко скопировать, поскольку они представляют собой не более чем программное обеспечение, и рассматривать все дела параллельно, а не последовательно; каждое дело получит своего собственного робосудью на то время, которое требуется. Наконец, судьи-люди не могут овладеть всей информацией, необходимой для всех возможных дел, от сложных патентных споров до убийств, и всеми последними достижениями криминалистики, робосудьи будут обладать практически неограниченной памятью и способностью к обучению» (Tegmark, 2018, p. 137).

Манифест киберутопизма

Суть киберутопизма (Morozov, 2012) почти полностью выражена в этой реконструкции, основанной на эмоциональной предпосылке в виде явных недостатков дорогостоящего, часто несовершенного и не всегда быстрого процесса правосудия. На смену человеческой необъективности должно прийти быстрое, экономичное и «определенное» решение машины. Эта предпосылка апеллирует к общепринятому мнению, однако она совершенно неполна и основана на очень опасном допущении: «юридические процессы можно абстрактно рассматривать как вычислительную задачу, где доказательства и законы являются входными данными, а на выходе мы получаем решение». Здесь кроется первая ошибка. Этот подход оказывает определенное влияние на обычного человека, но вряд ли может заменить собой полный и глубокий анализ проблемы.

Юридические процессы – это не вычисления, а размышления. Они касаются людей и фактов, которые никогда не бывают полностью идентичными. Математические правила нельзя применить к конкретным делам. Математические алгоритмы создаются именно для того, чтобы абстрагировать, обобщить, применить единую формулу к множеству различных данных.

Кроме того, не следует забывать об очевидном: если не право в целом, то судебная практика, безусловно, меняется, а «алгоритм» – нет. При ближайшем рассмотрении обнаруживается фундаментальная ошибка – теперь уже историческая, порожденная сходством между нейронными и вычислительными сетями. Человеческий мозг сравнивают с компьютером, что уже было опровергнуто теорией вычислений. Основатели теории вычислений, а в какой-то мере и творцы так называемого искусственного интеллекта, интуитивно понимали это на концептуальном уровне.

Таким образом, прежде чем рассуждать о предполагаемом (ошибочно) применении искусственного интеллекта и роботов-судей к праву и судопроизводству, стоит вернуться к некоторым соображениям о сходстве мозга и ком-

пьютера. Важно отметить, что киберутопизм, преследуя цель стимулировать рост вычислительных мощностей и распространение приложений ИИ, провозглашает основополагающую ценность «материально ничем не ограниченной памяти и способности к обучению», но при этом сознательно отбирает и распространяет лишь некоторые из гипотез основателей теории вычислений.

Это не голословное обвинение. Если так поступают разработчики нового искусственного интеллекта, то как мы можем верить в беспристрастность, нейтральность и «всеобщность» их творений? Вспомните об этом, когда в следующий раз услышите об ИИ, который «содержит в себе все», потому что в «очень большом» легко потерять что-то количественно незначительное, но содержательно и концептуально важное.

Результаты исследования

Короткая память киберутопистов о теории вычислений и ИИ

Alan Turing в работе “Proposed electronic calculator” («Проект электронного вычислителя») (Turing, 1946)¹ утверждал:

«Вероятно, самой серьезной проблемой для нас будет убедиться, что вычислитель делает то, что должен. Предположительно, можно выделить три вида ошибок.

- (1) *Постоянные неисправности, возникающие в проводах или деталях, например, размыкание конденсаторов.*
- (2) *Временные ошибки, вызванные помехами, шумом, достигшим неожиданно высокого уровня, необычными сочетаниями напряжений в некоторых точках цепи и др.*
- (3) *Ошибки, связанные с использованием неверных инструкций или с неверными представлениями о том, что должна делать схема».*

В ближайшее время мы планируем установить контролирующие схемы для выявления ошибок первого типа. Ошибки второго типа не должны возникать, если аппарат находится в исправном состоянии, однако, когда какой-либо компонент начинает давать сбои, его недостатки часто проявляются именно таким образом. Неправильные инструкции (ошибки третьего типа) часто проявляются в результате проверок, которые были заложены в эти же инструкции» (Turing, 1946, p. 16). «Есть три основные функции, которые должна выполнять проверка. Она должна устранять возможность ошибки, помогать диагностировать неисправности и внушать доверие. <...> Чтобы внушить доверие, проверка должна иметь какие-то видимые проявления. Конечно, всякий раз, когда проверка не пройдена, машина должна сообщать об этом» (Turing, 1946, p. 17).

Основоположник теории вычислений уверенно обозначает три типа ошибок. Третий тип, вероятно, свидетельствует о некоем технократическом утопизме: это идея, что делаться будет то, что должно быть сделано, как если бы это была научная потребность, которую он не принимает во внимание из экономических интересов (в соответствии с подходом и видением А. Turing’a). Но из задач проверки киберутопистов действительно волновала только одна: внушить доверие. Остальные воспринимались как неважные и часто упускались из виду.

В работе 1948 г. “Intelligent machinery: a report” («Интеллектуальные машины: доклад») Turing писал:

«Я предлагаю исследовать вопрос о том, возможно ли, чтобы машины демонстрировали разумное поведение.

- (а) *Нежелание признать, что у человечества могут быть конкуренты в области интеллекта.*
- (б) *Религиозное убеждение, что любая попытка построить такие машины – это своего рода гордыня Прометейя.*
- (в) *Очень ограниченный характер машин, которые использовались до недавнего времени (скажем, до 1940 г.). В результате возникло убеждение, что возможности машин всегда будут ограничены исключительно простыми, повторяющимися действиями.*
- (г) *Недавно теорема Геделя и связанные с ней результаты... показали, что если использовать машины для таких целей, как определение истинности или ложности математических теорем, и не принимать случайных неверных результатов, то конкретная машина в некоторых случаях вообще не сможет дать ответ».*

Если «машина может проявить интеллект, это следует рассматривать не иначе как отражение интеллекта ее создателя» (Copeland, 2004, p. 410; Turing, 1948, p. 2).

К этим, казалось бы, схематичным и упрощенным утверждениям Turing, однако, добавляет два соображения, очень важные для целей нашего анализа:

«Аргумент из теоремы Геделя и других теорем (возражение “г”) опирается, по сути, на условие, что машина не должна совершать ошибок. Но это не является требованием к интеллекту. <...> Утверждение “д” о том, что интеллект машины является лишь отражением интеллекта ее создателя, скорее, похоже на мнение, что заслуга в открытиях ученика должна принадлежать его учителю» (Copeland, 2004, p. 411; Turing, 1948, p. 3).

¹ Turing, A. M. et al. (1946). Proposed electronic calculator. Report, National. https://www.alanturing.net/turing_archive/archive/p/p01/P01-001.html

Таким образом, выходя далеко за рамки технологического аспекта и обращаясь к основам теории вычислений, Turing еще семьдесят лет назад предостерегал от двух грубых заблуждений, о которых сегодня, очевидно, забыли (как будто технический прогресс, увеличивая вычислительные мощности, избегает фундаментальных ошибок): то, что машина не должна совершать ошибок, не является критерием интеллекта (тем более искусственного) и что интеллект машин – это лишь отражение интеллекта их создателя (идея, используемая для успокоения). Наконец, общая тема, над которой необходимо подумать (человеческим интеллектом), такова:

«Степень, в которой мы считаем что-либо разумным, определяется как нашим собственным состоянием ума и подготовкой, так и свойствами рассматриваемого объекта» (Copeland, 2004, p. 431; Turing, 1948, p. 19).

В этом контексте заслуживает внимания и тот момент, который, по сути, стал провокационным вызовом (хотя вовсе не задумывался как таковой!) «проверке» машинного интеллекта. Alan Turing предвидел многие из самых серьезных возражений, которые могут быть выдвинуты против искусственного интеллекта. В своей фундаментальной статье *“Computing machinery and intelligence”* («Вычислительные машины и интеллект») (Turing, 1950) он перечисляет девять из них и цитирует отрывок из речи Geoffrey Jefferson:

«Пока машина не сможет написать сонет или сочинить концерт благодаря своим мыслям и эмоциям, а не случайному выпадению символов, т. е. не только написать, но и знать, что она это написала, можем ли мы согласиться с тем, что машина равна мозгу? Ни один механизм не может чувствовать (а не просто искусственно подавать сигналы, что легко сделать) удовольствие от своих успехов, горе, когда срываются его клапаны, не может согреться лестью, стать несчастным от своих ошибок, очарованным сексом, разгневанным или подавленным, если не получает желаемого» (Jefferson, 1949, p. 1110).

Все мы знаем основные правила и условия «игры Тьюринга», которую он придумал и описал в статье, процитированной выше. Менее известен один из вариантов игры, предложенный Turing’ом, который показывает механизм решения на чисто логической или почти логической основе. В этом варианте машина должна действовать совместно с человеком. Два участника заперты в разных комнатах и печатают свои ответы на ряд вопросов, с помощью которых допрашивающий пытается выяснить их пол. В контексте игр Turing’a важно помнить, что участник не обязан говорить правду.

Не следует забывать, что Turing никогда не обещал создать машину, абсолютно неотличимую от человека с любой точки зрения (именно за эту задачу взялись компании, заинтересованные в прибылях, полагая, что они могут это сделать). Он скорее стремился прояснить вопрос о том, могут ли машины мыслить. Он принимал как факт, что люди мыслят, а в качестве критерия того, может ли машина мыслить, предлагал построить такую машину, чтобы человек не мог отличить ее результаты от ментальных продуктов другого человека.

Другим авторитетом в области теории вычислений является Jeremy Bernstein (Bernstein, 1978; Bernstein, 1990), автор основополагающего труда *“Science Observed”* («Наблюдаемая наука») (Bernstein, 1982). Работа Bernstein’a актуальна по двум причинам. Во-первых, она «упорядочивает и суммирует» большую часть того, что было сделано в течение около тридцати лет в области теории вычислений и ИИ. Во-вторых, она рассматривает вопрос о сходстве между мозгом как разумом (интеллектом человека и животных) и вычислительной машиной (искусственным интеллектом), показывая огромные концептуальные различия задолго до известных открытий в нейробиологии.

Один из отцов теории вычислений Marvin Minsky (Minsky, 1965, p. 4; см. также Bernstein, 1990, p. 16) писал:

«Если человек хорошо понимает машину или программу, у него не возникает желания приписать ей ‘волеизъявление’. Если же человек понимает ее не так хорошо, он вынужден предоставить для объяснения неполную модель. Наши повседневные интуитивные модели высшей человеческой деятельности весьма неполны, и многие понятия в наших неформальных объяснениях не выдерживают пристального рассмотрения. Свободная воля или волевое усилие – одно из таких понятий: люди не в состоянии объяснить, чем оно отличается от стохастического каприза, но твердо уверены, что отличие есть» (Minsky, 1965, p. 4).

Bernstein продолжает эту мысль:

«Когда будут созданы разумные машины, мы не должны удивляться тому, что они окажутся такими же, как и человек, упрямыми и запутавшимися в своих убеждениях, касающихся проблемы соотношения разума и материи, свободы воли сознания и тому подобного. На самом деле все эти проблемы направлены на объяснение сложных взаимодействий между различными частями самодетели» (Bernstein, 1982, p. 12).

Он также писал о компьютере, созданном в 1951 г.:

«Случайные связи служат своего рода защитой от сбоев: если один из нейронов не работает, последствия будут не слишком серьезными... Я не думаю, что мы когда-нибудь сможем узнать, как устранить все ошибки в нашей машине... но это не имеет значения, потому что благодаря этим случайным связям она почти наверняка будет работать независимо от того, как она устроена» (Bernstein, 1990, p. 17).

По описанию Minsky, машина, построенная в Гарварде, была в основе своей скиннеровской. Ее действия без обратной связи были более или менее случайными, что ограничивало возможности для ее обучения; она так и не смогла создать план действий.

Здесь мы возвращаемся к параллелизму вычислительных сетей и нейросетей. Вопрос, на который пытались ответить McCulloch и Pitts (1943), заключается в том, на что способна нейронная сеть такого рода. Список их аксиом выглядит следующим образом:

- 1) активность нейрона – это процесс «все или ничего» (нейрон либо активен, либо нет);
 - 2) в любой момент времени для возбуждения нейрона необходимо, чтобы в течение латентного периода сложения возбудилось определенное количество синапсов;
 - 3) единственной значимой задержкой в нервной системе является синаптическая задержка;
 - 4) если активен какой-либо тормозной синапс, возбуждение нейрона в этот момент абсолютно исключено (теперь мы знаем, что они добавляются к остальным входным сигналам на синапсе);
 - 5) структура сети не меняется со временем (это требует переосмысления).
- Последняя аксиома предположительно означает, что эта идеальная сеть бессмертна.

Конец эпохи равенства между мозгом и компьютером

Через несколько лет после появления вышеупомянутой статьи создатель как минимум трех первых в истории суперкомпьютеров von Neumann (1958) написал, вдохновившись их результатами:

«Все, что может быть исчерпывающим и непротиворечивым образом описано, что может быть однозначно выражено словами, – все это фактически осуществимо с помощью соответствующей конечной нейронной сети» (Bernstein, 1990, p. 92).
«Подавляющее большинство современных исследователей ИИ считают, что им необходимо создавать все более сложные программы. В свете того, что было достигнуто к настоящему времени, эта деятельность оправдана, плодотворна, но она может также вводить в заблуждение» (Von Neumann, 2014, p. 135).

Следующее наблюдение было сделано F. Crick'ом (Crick, 1955) в статье, опубликованной в журнале *Scientific American*:

«Появление больших, быстрых и дешевых компьютеров дало нам представление о результатах, которые можно получить с помощью быстрых вычислений. К сожалению, аналогия между компьютером и мозгом, хотя и полезная в некоторых отношениях, может ввести в заблуждение. В компьютере информация обрабатывается очень быстро в последовательном режиме. В мозге скорость обработки гораздо ниже, но информация может обрабатываться параллельно по миллионам каналов. Операции современного калькулятора очень надежны, но, устранив одну или две из них, можно нарушить весь процесс. Напротив, нейроны мозга гораздо менее надежны, но исключение некоторых из них вряд ли внесет заметную разницу в результат. Калькулятор работает строго на основе двоичного кода. По-видимому, мозг использует менее точные способы коммуникации, но, с другой стороны, вполне вероятно, что он сложным образом регулирует количество и работу синапсов, чтобы адаптировать их функционирование к имеющемуся опыту. Поэтому неудивительно, что, хотя компьютер может точно и быстро выполнять длинные и сложные арифметические вычисления, в чем человек не силен, однако человек способен распознавать формы, что на сегодняшний день ни один компьютер даже отдаленно не может повторить» (Crick, 1979).

В 1966 г. von Neumann:

«...обсуждал общую конструкцию самовоспроизводящегося автомата. Он говорил, что в принципе можно создать мастерскую, которая при достаточном количестве времени и сырья может изготовить копию любой машины. В этой мастерской будет находиться станок В со следующими возможностями. Если дать ему объект X, то он исследует его, перечислит все его части и их соединения и таким образом получит описание объекта X. Используя это описание, станок В затем изготовит копию объекта X. Это очень близко к процессу самовоспроизводства, потому что можно дать станку В исследовать самого себя» (Neumann, 1966, pp. 82–83).

На конференции в Ванзе физик Freeman Dyson (Freeman, 1970) соотнес элементы машины von Neumann'a с их биологическими аналогами. Мастерская соответствует рибосомам, копировальная машина – ферментам РНК-полимеразы и ДНК-полимеразы, бригадир – это управляющие молекулы: репрессор и индуктор, а проект – РНК и ДНК. Von Neumann был первым, кто это понял.

Как создать систему, которая будет надежной, даже если операции, на которых она основана, не вполне надежны? Von Neumann понял, что ключ лежит в избыточности – наличии множества идентичных устройств с одинаковыми функциями. Он предположил, что центральная нервная система должна иметь избыточную организацию, чтобы функционировать с приемлемым уровнем ошибок. Он также понимал, что если универсальную машину, созданную Turing'ом, заставить воспроизводить саму себя, то она сможет эволюционировать. Если бы программа была изменена мутацией и, несмотря на это изменение, машина могла продолжать воспроизводиться, то возник бы модифицированный потомок.

По мнению исследователей F. J. Dyson (1985) и G. Dyson (2012), von Neumann был убежден, что сама возможность существования универсального автомата в конечном итоге делает мыслимой неограниченную биологическую эволюцию. В процессе эволюции простых организмов в более сложные нет необходимости каждый раз переделывать фундаментальный биохимический механизм, достаточно лишь модифицировать и расширять генетические инструкции. Von Neumann ясно выразил это в работе *“The Computer and the Brain”* («Компьютер и мозг») (von Neumann, 1958), но с оговоркой, которая должна была – но не сделала этого – закрепить симметрию и предполагаемое сходство между мозгом (интеллектом) и машиной (искусственным интеллектом):

«Нервная система – это вычислительная машина, которой удается выполнять свою чрезвычайно сложную работу на довольно низком уровне точности... ни одна известная вычислительная машина не может надежно и значимо работать на таком низком уровне точности» (von Neumann, 1958, pp. 75–76).

Киберутописты и вопрос «кто умнее»

Фраза «...ни одна известная вычислительная машина не может работать...» предопределила «гонку вооружений» в сфере искусственного интеллекта: чтобы опровергнуть рассуждения фон Неймана, одного из самых блестящих умов в истории, нужно было создать «неизвестную» машину. Она должна была называться искусственным интеллектом, должна была (как предписывал Turing) «внушать доверие», и по этой причине параллель с человеческим мозгом – вынужденная и фальсифицирующая научную реальность – была названа «безвредной», чтобы «познакомить» нас с новым творением.

Другой величайший ум XX в. также должен был быть повержен одним ударом: Alan Turing, играющий в свою игру (первую, а не вторую, которая была бы слишком сложной).

Вызов киберутопизму был брошен.

Киберутопизм права: как роботы-судьи будут судить человека

Существует множество различных точек зрения на связь между искусственным интеллектом и уголовным правосудием. Эта отрасль права вызывает наибольшее количество споров и сравнений; она также требует важнейших навыков регулирования и защиты прав, которые не входят в компетенцию других отраслей знания. Дебаты ведутся вокруг вопросов неприкосновенности частной жизни и таких все более узких областей, как защита нейронных прав, авторское право, защита конфиденциальности персональных данных и биологических метаданных вплоть до области эмоционального ИИ.

Оставив пока в стороне эти аспекты, которые тем не менее очень важны, выделим некоторые из наиболее значимых позиций в сфере уголовного правосудия.

Как считает P. M. Morhat, искусственный интеллект может выполнять такие функции в отправлении правосудия, как рутинная работа судьи, включая контроль за ходом судебного заседания, обработка информации, оформление документов, экспертно-аналитическое, лингвистическое и организационное обеспечение судопроизводства (Morhat, 2018, pp. 25–27). Более того, существующая практика использования ИИ в судопроизводстве показала, что невозможно разграничить действия судьи-человека и ИИ, особенно в сложных делах; напротив, они должны сотрудничать, чтобы судья-человек мог контролировать ИИ. Ключевым вопросом является анализ алгоритма коммуникации между искусственным интеллектом и человеком (Buocz, 2018).

Исследователи Andreev, Laptev и Chucha (2020, p. 31) отмечают, что работа ИИ в судебном процессе не может быть изолирована от действий судьи-человека. Получить независимые «решения машины» (судебные акты) пока невозможно. Оценка человека-судьи необходима в таких вопросах, как юридическая квалификация и др. Решение судьи принимается в ходе самого процесса. Однако машина определяет приговор без учета взаимодействия подсудимого с судьей (судом). Она не оценивает вербальные и невербальные средства общения (тон голоса, мимику, жесты, позу). Эти и другие наблюдения формируют психологический портрет подсудимого и способствуют тому, чтобы объем исправительного воздействия был адекватен уголовному делу (Alikperov, 2018, p. 21).

Эти вопросы рассматриваются в концептуально важной части труда *“Life 3.0”* («Жизнь 3.0») М. Tegmark’a, где говорится о применении ИИ к праву. Однако его рассуждения следует с осторожностью применять к другим сферам человеческой деятельности: медицине, вождению автомобилей, научным исследованиям или преподаванию. Некоторые сложности нельзя устранить, используя машину вместо человека. В сфере закона и права ситуация выглядит следующим образом:

«Какие первые ассоциации приходят вам в голову, когда вы думаете о судебной системе в вашей стране? Если это длительная волокита, высокие расходы и иногда несправедливость, то вы не одиноки. Разве не было бы замечательно, если бы в первую очередь вы подумали об эффективности и справедливости?» (Tegmark, 2018, p. 136).

«Поскольку юридический процесс можно абстрактно рассматривать как вычислительную задачу, в которой вводится информация о доказательствах и законах и выдается решение, некоторые ученые мечтают полностью автоматизировать его с помощью робосудей. Это системы искусственного интеллекта, которые будут неустанно применять высокие юридические стандарты в каждом решении, не поддаваясь человеческим ошибкам, таким как предвзятость, усталость или отсутствие новейших знаний» (Tegmark, 2018, p. 136).

«Впервые в истории робосудьи в принципе могли бы обеспечить всеобщее равенство перед законом. <...> Их можно запрограммировать так, чтобы все они были одинаковыми и относились ко всем одинаково, применяя закон на основе прозрачности и настоящей объективности... Робосудьи могли бы также устранить те человеческие предубеждения, которые являются случайными, а не преднамеренными. Еще один недостаток человеческих судей заключается в том, что им не хватает времени на изучение всех деталей дела. В отличие от них робосудей можно легко скопировать, поскольку они представляют собой не более чем программное обеспечение, и рассматривать все дела параллельно, а не последовательно; каждое дело получит своего собственного робосудью на то время, которое требуется. Наконец, судьи-люди не могут овладеть всей информацией, необходимой для всех возможных дел, от сложных патентных споров до убийств, и всеми последними достижениями криминалистики, робосудьи будут обладать практически неограниченной памятью и способностью к обучению» (Tegmark, 2018, p. 137).

Что такое (человеческий) интеллект и почему ИИ – не интеллект

В своем выступлении на тему «Алгоритмы и искусственный интеллект. Люди совершают ошибки, именно поэтому они лучше машин» Ginevra Cerrina Feroni (2020), вице-президент Управления по защите персональных данных Италии, заявила:

«Мы должны пересмотреть понятие 'интеллект'. Это совсем не очевидная операция, потому что это понятие может иметь множество различных значений: способность к логическому мышлению, понимание, планирование, самосознание, творчество, решение проблем, обучение...»

Однако она добавляет:

«Мы можем принять широкое определение интеллекта, как это предлагает Max Tegmark (*Life 3.0: Being human in the age of artificial intelligence: Vintage*): интеллект – это 'способность достигать сложных целей' или 'способность приобретать и применять знания и компетенции'» (Feroni, 2020).

Возможно, настало время рассмотреть первый вопрос.

Определение интеллекта

Даже в случае общего понятия его определение нельзя выбрать произвольно, или путем максимального обобщения, или по критерию «наибольшей объяснительной эффективности». Дело в том, что большая часть представителей киберутопизма не инженеры или программисты, не математики или специалисты в науках об информации, что было бы логично, а мыслители и популяризаторы, которых щедро финансируют для того, чтобы создать «видимость достоверности» и ощущение близости к «чужой» реальности.

В бестселлере "*Machine, platform, crowd: Harnessing our digital future*" («Машина, платформа, толпа: наше цифровое будущее») (McAfee & Brynjolfsson, 2017) термин «защита частной жизни» не используется ни разу на 350 страницах! Однако текст на задней обложке гласит:

«Творчество, которое определяет наше будущее в цифровом мире, приходит с периферии. Новое использование машин с искусственным интеллектом меняет экономику и наш образ жизни. Два великих светила показывают нам, как жить в условиях революции» (McAfee & Brynjolfsson, 2017).

При этом подразумеваются соответственно директор и главный научный сотрудник Центра цифрового бизнеса Массачусетского технологического института; их исследования финансируются Илоном Маском на сумму около 10 млн долл. Было бы уместно выяснить определение интеллекта у нейробиологов, психологов и антропологов (возможно, из числа тех, кого не финансируют компании или магнаты новой экономики).

Существуют простые определения интеллекта, например, способность адаптироваться к окружающей среде. Некоторые из них более всеобъемлющие и сложные, с большим количеством элементов. Хорошим примером является современная концепция окружающей среды, где она определяется не просто как место проживания, а как активный элемент, способный изменять тех, кто в ней живет. Аналогичным образом, вероятно, и интеллекту человека невозможно дать одно-единственное определение.

Мы также можем обнаружить, что единого интеллекта не существует.

Профессор Гарвардского университета Howard Gardner в своей авторитетной книге *“Formae mentis. Essay on the plurality of intelligence”* («Структура разума: теория множественного интеллекта») (Gardner, 1983, см. также комментарии Di Salvo (2024)) предложил теорию множественного интеллекта. Он опроверг концепцию интеллекта как единого фактора, измеряемого с помощью коэффициента интеллекта (*IQ*), и дал более динамичное определение, выделив субфакторы, дифференцированные друг от друга.

В ходе эмпирических исследований, подкрепленных множеством источников в области нейропсихологических патологий, он выделил по меньшей мере семь различных типов интеллекта: лингвистический, музыкальный, логико-математический, пространственный, телесно-кинестетический, личностный и межличностный. Gardner также предположил, что эти формы (или специализации) человеческого интеллекта не являются чем-то исключительным, а представляют собой уникальную личностную комбинацию каждого индивидуума. По утверждению автора, в одном и том же человеке присутствуют все упомянутые типы интеллекта (а также другие, которые еще предстоит должным образом классифицировать), но с разной степенью качественного и количественного развития. Это сочетание типов интеллекта с абсолютно разным процентным соотношением, уникальное для каждого человека, определяет жизнь и адаптацию как отдельных людей и групп, так и общества и вида в целом.

ИИ, которому можно доверять

В апреле 2019 г. Европейская комиссия разработала «Этические принципы для надежного ИИ», в которых говорится, что «процессы должны быть прозрачными, возможности и назначение систем ИИ должны открыто сообщаться, а решения, по мере возможности, объясняться тем, кого они прямо или косвенно затрагивают. Без такой информации решение не может быть должным образом оспорено» (European Commission, 2019, p. 13).

Известный исследователь G. Cerrina Feroni отмечает:

«К сожалению, приходится признать, что не всегда можно объяснить, почему та или иная модель привела к определенному результату или решению. Это происходит потому, что в большинстве своем алгоритмы являются непостижимыми системами, своего рода ‘черными ящиками’, если воспользоваться определением Frank Pasquale [(Pasquale, 2015)]. Черные ящики изучают характеристики пользователей, составляют профили, выносят суждения, предлагают решения. Все это, однако, без объяснения того, как они пришли к тому или иному выводу» (Feroni, 2020).

Принцип понятности предполагает, что алгоритмы, особенно прогностические, основаны на логике, т. е. на прослеживаемой и проверяемой последовательности аргументов. Проблема в том, что большинство алгоритмов нового поколения не ограничиваются детерминированным выведением следствий из заданных аксиом, но благодаря системам автоматического обучения сами вырабатывают критерии, которые во многих случаях непонятны даже программистам (Simoncini, 2019).

G. Cerrina Feroni далее поясняет:

«Правовые нормы в области искусственного интеллекта ставят вопрос о принципе ответственности, который неразрывно связан с рамками разумности в человеческой деятельности. Кто должен нести юридическую ответственность за решения и действия искусственного интеллекта, например, автономного автомобиля, робота-хирурга или вооруженного беспилотника? Вопрос о юридической ответственности в случае с искусственным интеллектом, который обучается автономно, является огромным и до сих пор не решенным. <...>

В гонке машин, все лучше запрограммированных на достижение целей с максимальной эффективностью, никто не задумывается об ошибках, помехах, вариациях...

Поэтому необходимо изменить точку зрения. Осознавая свое физиологическое несовершенство, мы должны заявить о праве на ошибку. В противном случае мы рискуем оказаться жертвами тоталитаризма эффективности, возможно, устойчивой с экономической и экологической точек зрения, но не человеческой. Напротив, бесчеловечной. Эта эффективность может зайти далеко, вплоть до таких парадоксальных сценариев, как самоуничтожение, рассчитанное с точки зрения коллективной выгоды» (Feroni, 2023).

Предполагаемый технологический нейтралитет

Прежде чем вернуться к исходной гипотезе, т. е. к предполагаемой гиперэффективности робота-судьи, уместно рассмотреть еще одно допущение, которое киберутописты считают само собой разумеющимся и громко декларируют с единственной целью – представить ИИ надежным и безопасным. Речь идет о предполагаемой «нейтральности» технологий.

В своей работе *“Artificial intelligence and the risks of a ‘disruption’ of legal regulation”* («Искусственный интеллект и риски “нарушения” правового регулирования») Giuseppe Mobilio раскрывает эту тему:

«Право должно преодолеть первое заблуждение, которое подпитывает идеологическое недоверие к правовому регулированию, а именно идею о том, что технология ‘нейтральна’, т. е. имеет чисто инструментальную ценность и поэтому не нуждается в регулировании, хотя бы в самом общем (Greenberg, Rethinking Technology Neutrality, in Minnesota Law Review, 2016)» (Mobilio, 2020, p. 406).

Столкнувшись с таким предположением, можно вспомнить С. Schmitt'a и его книгу "The concept of 'political'" («Понятие политического») (Schmitt, 1927). Говоря о «вере в технику», он писал: «Техника всегда только инструмент и оружие, и именно потому, что она служит всем, она не нейтральна... Это оружие, которое можно повернуть в сторону любой цели и которое остается культурно и нравственно слепым».

Tegmark писал, что роботы-судьи «могут быть запрограммированы так, чтобы быть одинаковыми и относиться ко всем одинаково, применяя закон прозрачно и действительно объективно. Они также могли бы устранять случайные, непреднамеренные проявления предвзятости человека». Отвечая на это предположение, Mobilio уточняет:

«ИИ строится на алгоритмах. И нужно учитывать, что не существует алгоритмов, которые просто нейтрально отражают реальность; скорее, чтобы получить ожидаемые результаты, они предлагают собственное представление поставленных проблем, и результат определяется выбранными переменными, формулами классификации, весами отдельных вводимых параметров и применяемыми процедурами. Кроме того, решения часто принимаются алгоритмами, предназначенными не только для использования данных, непосредственно представляющих определенное явление, но и для обработки, прежде всего, опосредованных или косвенных (прокси) данных, которые могут снизить точность решений и еще больше исказить представление о реальности.

Если и этого недостаточно, всегда существует риск, что данные, на которых изначально обучались алгоритмы, отражают проявления неравенства, отвержения или других видов дискриминации, присутствующих в обществе. Это так называемые предвзятости, или 'искажения', которые 'систематически и несправедливо дискриминируют одних людей или группы в пользу других' и которые, в свою очередь, выливаются в дискриминацию получателей или заинтересованных сторон алгоритмических решений. Кроме того, каждая алгоритмическая модель разрабатывается на основе целей, выбранных ее создателями, и, следовательно, является идеологически ориентированной. Эта ориентация часто направлена на эффективность и прибыли тех, кто использует эти инструменты, а не на справедливость, равенство или благо общества» (Mobilio, 2020, pp. 406–407).

Вычислительный не-интеллект, или дефект искусственности

Возвращаясь к модели интеллекта Gardner'a, мы можем сразу же заметить двусмысленность термина «искусственный интеллект» и попытаться сделать его чуть менее «знакомым», а значит, приемлемым для человеческого сообщества.

Gardner выделил (как минимум) семь различных типов интеллекта: лингвистический, музыкальный, логико-математический, пространственный, телесно-кинестетический, личностный и межличностный. Из этих типов машина и ее алгоритмы, конечно же, не могут обладать лингвистическим и музыкальным творческими типами интеллекта, но могут обладать чисто исполнительскими. Это связано с тем, что «художественное» творчество требует взаимодействия интеллекта со средой, во-первых, движения (рецептивного взаимодействия) и, во-вторых, ценностного опосредования чувств (которые являются квалификаторами опыта).

ИИ также может обладать пространственным, телесно-кинестетическим, чисто вычислительным интеллектом: он может получать исходные данные из среды и обрабатывать их, управлять опытом в своей памяти, но не может выйти за пределы исходных данных в область ценностного опыта.

Машина полностью (онтологически) лишена личного и межличностного интеллекта, за исключением чисто самодиагностических (механическая математика для собственной диагностики) и релятивистских (механические вычисления во взаимосвязи – но не в отношениях – с «другими») задач. Она, несомненно, обладает логико-математическим интеллектом (в той мере, в какой эффективны и оптимизированы ее логико-математические вычислительные способности), но, конечно, не творческо-изобразительно-абстрактным (проще говоря, не может «изобрести или открыть» что-либо, начиная с категоризации и абстрагирования реальности).

Робот-судья

Робот, прекрасно умеющий играть в шахматы и решать сложнейшие уравнения, но лишенный лингвистического, музыкального, личностного и межличностного интеллекта, сопоставимого с человеческим, и имеющий пространственный, телесно-кинестетический интеллект, развитый лишь с технической и технологической точек зрения, – такой робот-судья предназначен для того, чтобы судить человека. Судить того, чья идентичность основана именно на сочетании всех типов интеллекта – лингвистического, музыкального, логико-математического, пространственного, телесно-кинестетического, личностного и межличностного, причем это сочетание субъективно-индивидуально, в нем присутствуют все эти типы, по-разному модулированные (качественно и количественно), развившиеся не только в среде, но и из среды, с которой машина не имела никаких значимых отношений.

Очевидно, что каждый судья при вынесении каждого отдельного решения, с учетом фактов и обстоятельств, а также людей, стоящих перед ним, будет судить по-разному (даже если эти различия невелики). Судье требуется время, чтобы рассмотреть все обстоятельства и принять решение о наказании: свобода человека не должна решаться за секунды. Поскольку суждение – это не «вычисление», а «размышление», его краеугольным камнем остается то, что

судья – хотя бы по основным биологическим и неврологическим параметрам – должен быть «равен» подсудимому. При всех рисках неточности и неидеальности человеческих суждений, они остаются, безусловно, предпочтительнее любых математико-механистических решений.

ТРИ ЗАМЕТКИ НА ПОЛЯХ

О книге «Жизнь 3.0 – быть человеком в эпоху искусственного интеллекта» Макса Тегмарка (“Life 3.0 – Being human in the era of artificial intelligence” by Max Tegmark)

1. О сознании машин

«PNI. Путешествие от мозга к душе» – так называется книга Giulio Tononi (Tononi, 2012). PNI придает теории сознания форму нарратива (чего никогда не сможет сделать *ChatGPT*). Oliver Sacks дал лучшее резюме этой работе, сказав: «Giulio Tononi придал телесность и духовность последним достижениям нейробиологии и превратил их в прекрасный роман».

Tononi писал:

«Наука дала нам великую силу, способность быстро двигаться, вырабатывать огромное количество энергии, создавать новые кристаллы и новые металлы, мгновенно передавать слова и изображения через всю планету. Но нельзя игнорировать то, о чем говорят нам эти открытия: мы навсегда заключены на задворках Вселенной и для этой Вселенной наша жизнь – всего лишь мгновение, лишь точка в пространстве» (Tononi, 2012).

В конце своей книги Tegmark вводит элемент сознания в вычислительные машины через развитие ИИ, приступая к рассмотрению теорий о сознании; однако он упомянул лишь некоторые из них, описав их в самом общем виде, чтобы работа была более информативной. Однако эта тема заслуживает большего внимания.

Верной представляется следующая мысль Tononi:

«Кажется, не следует делать вывод, что слабая реакция говорит о слабом сознании и, наоборот, что умные ответы обязательно подразумевают развитое сознание». По его мнению, «машина может давать умные ответы, но не иметь сознания, потому что она состоит из отдельных модулей... Поэтому, хотя она может иногда обманывать нас своими ответами, ей не хватает контекста и понимания, которые обеспечивает лишь сознание».

«В мозге есть мощные механизмы, и сознание мудро использует их все. Одни предназначены для понимания свойств нашего мира, другие – для действий. Самыми тонкими являются механизмы, которые вычисляют, планируют и советуют, а также те, что хранят память о событиях – хотя все они являются лишь инструментами, а не частью самого сознания» (Tononi, 2012).

Как мы уже говорили, и о чем напоминал Giulio Tononi, Turing, первым начавший дебаты об искусственном интеллекте, предложил, как проверить, может ли машина думать и, возможно, продемонстрировать сознание. Данный тест сегодня носит его имя. Эксперимент заключается в следующем: интеллектуальная машина и человек помещаются в одну комнату и общаются по телетайпу с человеком, находящимся в другой комнате. Если последний не сможет отличить ответы машины от ответов человека, то машина может считаться разумной.

В продолжение этой темы философ John Searle утверждал, что машина, выполняющая набор инструкций, не мыслит, не обладает сознанием и не понимает смыслов, даже если она достаточно адекватно отвечает на всевозможные вопросы и выполняет тест Turing’a. В работе «Разум, мозг и программы» (Searle, 1980) он предложил эксперимент: представьте, что вы заперты в комнате, имея набор карточек с китайскими иероглифами и инструкцию о том, какие карточки передавать вовне в ответ на входящие карточки; вы сможете выполнить программу, не понимая ни слова по-китайски. Этот аргумент подтверждает тезис о невозможности извлечения смысла (семантики) из механизмов (синтаксиса), но он также подтверждает, что сознание не может быть выведено из вычислений.

Gregory Bateson в работе “Steps to an ecology of mind” («Шаги к экологии разума») (Bateson, 1972) определил информацию как «разницу, которая делает разницу». Создателем теории информации является инженер и математик Cloud Shannon (Shannon, 1950). В своей работе он говорит о том, что информация – это число и что оно отделено от смысла. Аналогично тому, как расцвет науки произошел после того, как Galileo отделил наблюдателя от природы, так и коммуникации и хранение данных начали бурно развиваться после того, как Shannon разделил смысл и информацию.

Давняя традиция атомизма утверждает, что все можно объяснить, исходя из простых элементов и их взаимодействий. Является ли изображение просто набором точек? Это именно так: камера видит изображение как набор точек, поэтому на самом деле она его не видит.

Tononi продолжает:

«Следует подчеркнуть три момента. Во-первых, мы не одиноки. Существует социальный аспект сознания, который развивает его, расширяет, придает ему ценность и самореализацию. Во-вторых, все мы – люди, и все мы – жизнь. Человек может и должен отождествлять себя с другими живыми существами, это лишь вопрос степени, и в этом смысле его сознание бессмертно. В-третьих, тем не менее, каждый человек особенный и потому драгоценный» (Tononi, 2012).

2. О теории утопической киберзанятости и оплаты труда

Киберутописты настаивают, что новые технологии, в частности искусственный интеллект, создают и будут создавать все новые и новые механизмы для улучшения нашей жизни, сокращения или упразднения тяжелой, опасной, изнурительной работы, обеспечивая нам безбедную жизнь без усилий. В представлении киберутопистов мир состоит из менеджеров, управляющих всем на свете, которые работают мало (часто очень мало), а богатства, распределяемые в виде зарплаты, волшебным образом приумножаются (в гораздо больших масштабах, чем хлеба и рыбы). Само собой разумеется, что благодаря ИИ в мире больше не будет голода, а болезни если не исчезнут, то резко сократятся (и почти все они будут излечимы). Люди будут иметь «цели без работы» и «доход без работы».

В книге Tegmark'a "Near future" («Ближайшее будущее») (глава 3, называемая, разумеется, «Ближайшее будущее») этот удивительный мир описывается на десятке страниц в разделе «Занятость и вознаграждение», где также даются ценные советы о том, какие рекомендации по карьере следует давать нашим детям, чтобы быть уверенными, что завтра они окажутся «в нужном месте с правильной подготовкой». Очевидно, что нам показывают (ближайшее) будущее, которое не требует источников, исследований и документации. Дело в том, что, даже когда документация существует и ситуация кажется предсказуемой и почти само собой разумеющейся, история идет своим путем: алгоритмы не делают историю.

В 1917 г. все бы поставили на то, что если коммунистическая, или социалистическая, революция и произойдет, то в промышленно развитой Германии, а уж никак не в отсталой и, по сути, крестьянско-феодальной России. Никто бы не поверил, что в кайзеровской Германии возможна демократия. Смысл пророчества почти банален: если оно сбудется, то мы были правы, если не сбудется, то это потому, что государства увлеклись регулированием и люди «недостаточно верили». На самом деле все говорит о том, что антиутопический мир, подобный тому, что описан в "Brave New World" («Дивном новом мире») Aldus Huxley (Huxley, 1932), – это «ближайшее будущее», пророчески гораздо более вероятное в утопическом сценарии кибернетического ИИ по сравнению со сценарием Tegmark и McAfee.

В подтверждение этого пророчества на десятке страниц и еще более длинных пророчеств в тексте МакАфи можно обратиться к книге «Эра капитализма наблюдения: Борьба за человеческое будущее на новой границе власти» Shoshana Zuboff (Zuboff, 2019), американского социолога и эссеиста, получившей докторскую степень по социальной психологии в Гарвардском университете и степень по философии в Чикагском университете. Если для подтверждения идеи вам нужны академические титулы, то вы можете прочитать 630 страниц данных, конкретной информации и реальных фактов, собранных в книге. Однако не стоит забывать и о «нобелевской болезни», т. е. о всеобщем когнитивном предубеждении, которое называется «апелляция к авторитету» в поддержку своей теории.

В своей статье «Нобелевская болезнь: Когда интеллект не защищает от иррациональности» Candice Basterfield и соавт. (Basterfield et al., 2020) рассказывают о многочисленных случаях, когда мнение научных авторитетов, даже знаменитых нобелевских лауреатов, вне их собственной области специализации оказалось неверным. Эти мнения варьировались от евгеники до расовых заболеваний, от гомеопатии до астрологии, от СПИДа и ВИЧ до аутизма! И в этом случае немного здравого здравого смысла и «гуманного» использования человеческого интеллекта могут помочь не поддаваться на легкие и удобные утопические пророчества-обещания, которые – в данном случае явно – не бескорыстны, а обоснованы огромными экономическими интересами.

3. Голодный судья более жесток

Tegmark пишет, что робосудьи «могли бы также устранить те человеческие предубеждения, которые являются случайными, а не преднамеренными» (Tegmark, 2018, p. 137). Он подтверждает это следующей цитатой:

«Например, в 2012 г. было проведено неоднозначное исследование с участием израильских судей; утверждалось, что они выносят значительно более суровые вердикты, когда голодны: если сразу после завтрака они отклоняли около 35 % дел об условно-досрочном освобождении, то непосредственно перед обедом – более 85 %» (Tegmark, 2018, p. 137).

Прочитанное исследование (Danziger et al., 2011a) было раскритиковано как «некорректное» в работе Weinshall-Margel & Shapard (2011), хотя в ответ Denzinger и соавт. (Danziger et al., 2011b) продолжали настаивать на своих выводах.

Возможно, эти статистические данные показывают, что лучше представлять перед судом в 9 часов утра, а не в 12, и что полезно будет предложить судье круассан; но представляется несколько натянутым считать их доказательством того, что человека – который должен сам себя кормить – нужно заменить компьютером, подключенным к электрическому кабелю.

Статистические данные показывают, что человека лучше заменить компьютером, но говорить о том, что «роботы-судьи могли бы также устранить те человеческие предубеждения, которые являются случайными, а не преднамеренными», видимо, пока преждевременно.

Чтобы быть менее полемичными и придерживаться более научного подхода (в том числе и в интересах нашего будущего судьи-человека), мы должны принять во внимание – если мы считаем представленное исследование и выводы из него обоснованными, – что так называемое состояние *hangry* (сочетание английских слов *angry* – «сердитый» и *hungry* – «голодный») – это сложная эмоциональная реакция, вызванная сочетанием биологических особенностей, сигналов окружающей среды и личности человека. Агрессивная реакция в условиях длительного отсутствия пищи возникает не сразу, и есть два фундаментальных критерия, позволяющих определить, приведет ли оно к возникновению негативных эмоций: контекст и самосознание. Аппетит и настроение в целом очень тесно связаны между собой².

Возвращаясь к израильскому исследованию, следует сказать, что зачастую более простые дела об условном освобождении ставятся на начало дня, чтобы иметь больше времени для работы со сложными случаями. Поэтому вполне вероятно, что в более сложных делах судья принимает решение отказать в условно-досрочном освобождении, независимо от гликемического или серотонинового индекса. Возможно также, что после четырех-пяти часов умственной работы и напряженного и непрерывного внимания (как мы надеемся) общее состояние судьи подсказывало «в случае сомнений» не выпускать потенциального убийцу на свободу.

Тем не менее в каждом исследовании есть рациональное зерно. В данном случае оно касается не только судей, но и врачей, студентов и всех, кто занимается деятельностью, требующей внимания, концентрации и интеллектуальных усилий: перекусывайте каждые 2–3 часа, чтобы снизить гликемический индекс, лучше пищей, содержащей простые сахара и протеины, придерживайтесь здорового и сбалансированного питания в сочетании с физической активностью. Если вы чувствуете, что утратили равновесие, ощущаете усталость, голод, грусть и депрессию, избегайте принимать важные решения в этот момент. Это тоже интеллект, и в нем определенно нет ничего искусственного.

Заключения и не-заключения

Я заимствую это название из книги *“The lie of power”* («Ложь власти») Giulio Maria Chiodi (Chiodi, 1979), который был моим преподавателем философии политики. Здесь мы не обыгрываем название, чтобы поговорить о «лжи» в отношении определения «интеллекта», приписываемого алгоритмам, на основе которых вычислительные сети функционируют и перемалывают данные. В этом исследовании также не обсуждаются инструменты, с помощью которых компании новой экономики отвлекают внимание от самих себя, занимая позицию, основанную на двух постулатах: во-первых, мы саморегулируемы, потому что государственное регулирование неэффективно и вредно; во-вторых, все, что мы делаем, мы делаем для того, чтобы улучшить вашу жизнь. Именно на этих предпосылках строится теория киберутопизма.

Я заимствую название – и только его – потому, что (и это правило выходит за рамки конкретной темы упомянутой книги) критика должна заставлять нас размышлять. Она никогда не ставит точку; ее цель – вызвать дискуссию и стимулировать размышления, интеллектуально честные, как можно надеяться. Учитывая, что эта рефлексия «человеческая» и использует «человеческий интеллект» в качестве инструмента, она, по самой своей сути, содержит ошибки и упущения. Хочется надеяться, что благодаря этим ошибкам мы сможем со временем улучшить себя, как это всегда делал *homo sapiens*, не полагаясь на простое, казалось бы, решение – как предлагает технократический утопизм, делегировать другим (в данном случае другому, запрограммированному другими с неизвестными критериями) право «рассуждать и принимать решения за нас».

² Прием пищи – это сложное поведение; оно не просто означает питание, но подразумевает тесную связь между физиологическими, психологическими, социальными, семейными и символическими аспектами. Взаимодействие всех этих факторов порождает наши отношения с едой, которые каждый из нас выстраивает с детства и которые определяются ощущением голода и последующим чувством сытости. В основе этого механизма лежит нейромедиатор серотонин, вырабатываемый в основном в желудочно-кишечном тракте, а также в головном мозге:

P. S.

Сказанное относится к области права и правосудия. Но это касается и всех остальных сфер человеческой жизни: развлечений, культуры, искусства, здравоохранения, управления транспортными средствами, поездами, самолетами, кораблями. Выбор, какой фильм смотреть или как реагировать на турбулентность в полете, может быть обоснованно подкреплен сложными технологическими инструментами прогнозирования, но принятие решения, действие должно оставаться за человеком. Так я думаю, с помощью своего мозга, моей личной комбинации типов интеллекта, которые работают, как и ваши, посредством химических процессов и электричества, запрограммированные в архитектуре нашей ДНК, наполовину от каждого родителя.

Я наполнен содержанием из всего, что меня окружает, из моего опыта, из того, что я читал – в той последовательности, в которой читал, из фильмов, которые я видел, из музыки, которую слушал; я соткан из людей, которых встречал, из улыбок, которые получал, и из слез, которые пролил; из пейзажей, рассветов и закатов, которые видел, и ароматов цветов, которые вдыхал. Из тел, которых я касался, из взглядов, которые встречал, и глаз, в которых отжался. Из животных и деревьев, к которым я притрагивался, и слов, которые слышал.

Все это – не просто «содержание памяти», а человеческие воспоминания, которые имеют дополнительную «ценность», называемую эмоциями и чувствами, – делает меня тем, кто я есть.

И всего этого никогда не будет у вычислительной машины, которая со своим не-интеллектом даже не знает, что все это значит.

Но вы знаете, потому что вы – люди, млекопитающие, представители вида *homo sapiens*, как и я.

Давайте оставаться людьми.

– в мозге он регулирует настроение, качество сна, температуру тела, сексуальность и аппетит. Собственно, недостаток серотонина вызывает депрессию, обсессивно-компульсивные расстройства, тревогу, мигрени, нервный голод и булимию, преждевременную эякуляцию у мужчин;

– в кишечнике серотонин контролирует моторику кишечника, гигиену и самочувствие. Его недостаток приводит к хроническим запорам, нарушению пищеварения и замедлению кишечного транзита.

Уже доказано, что мозг и кишечник связаны между собой; поэтому у людей, страдающих депрессией, нарушена работа кишечника, а при нервном голоде также наблюдается изменение обмена серотонина.

Предшественником серотонина является триптофан – аминокислота, которая не синтезируется в организме человека, но присутствует в животных и растительных белках.

Проблема в том, что триптофан попадает в клетки человека только после других аминокислот. До триптофана в клетки, особенно кишечника и нервной системы, попадают другие аминокислоты: изолейцин, лейцин, тирозин, валин, метионин, фенилаланин. В результате снижается выработка серотонина, а также мелатонина, который является производным от первого. Вследствие этого изменяется гормональный баланс, и в нас нарастает так называемый голод по сладкому.

На самом деле продукты с большим содержанием сахара повышают уровень серотонина в центральной нервной системе. При потреблении сахаров вырабатывается гормон инсулин, который облегчает поступление в клетки аминокислот, за исключением триптофана.

В результате триптофан продолжает циркулировать в крови и усваиваться центральной нервной системой. Это объясняет, почему сладкие продукты способны повышать уровень серотонина, а значит, и хорошее настроение.

С точки зрения диетологии, необходимо употреблять низкокалорийные продукты, богатые триптофаном, но бедные другими аминокислотами. Этой характеристике частично удовлетворяют лишь некоторые фрукты, такие как папайя, бананы и финики. Почему бы не ввести в рацион больше мяса и его производных, очевидно, более богатых этой аминокислотой? Именно потому, что мясо богато различными аминокислотами, которые преобладают над триптофаном. Средние значения представляют рыба и яйца.

Еще один способ повысить уровень серотонина – заняться физической активностью. Мышцы в основном потребляют аминокислоты с разветвленной цепью, которые используются для построения мышц, и сохраняют триптофан, который таким образом остается доступным для нервных клеток.

Следует также помнить, что триптофан не превращается в серотонин без правильного потребления витаминов В3, В6 и С, которые содержатся в большей концентрации в следующих продуктах:

- витамин В3: пшеница, ячмень, бобовые, томаты, молоко, сыр, рыба, морковь, картофель;
- витамин В6: молоко, рыба, крупы, картофель, сыр, шпинат, фасоль, морковь;
- витамин С: свежие фрукты и овощи, особенно цитрусовые, киви, перец, брокколи.

В этом контексте оптимальна диета с большим содержанием яиц, персиков и овощей. Наконец, последние исследования показали, что достаточное присутствие жиров омега-3 в мозге оказывает антидепрессивное и стимулирующее воздействие на производство нейронов, ведь нейроны состоят главным образом из жиров.

Диабет – это заболевание, зависящее от эмоций и настроения, поскольку существует тесная функциональная связь между эндокринной и лимбической системами, задним отделом мозга, или «аффективным мозгом».

Так, внутреннее спокойствие и эмоциональное равновесие стабилизируют уровень сахара в крови и облегчают контроль над диабетом. Напротив, беспокойство или тревога обычно сопровождаются выбросом таких гормонов, как адреналин и кортизол, которые препятствуют действию гормона инсулина; последний, не имея возможности свободно передавать сахар из крови в клетки, приводит к состоянию гипергликемии.

References / Список литературы

- Alikperov, H. D. (2018). Electronic system for determining the optimal punishment (formulation of the problem). *Kriminologiya: vchera, segodnya, zavtra – Criminology: Yesterday, Today, Tomorrow*, 4, 51.
- Andreev, V. K., Laptev, V. A., & Chucha, S. Yu. (2020). Artificial intelligence in the system of electronic justice by consideration of corporate disputes. *Vestnik of Saint Petersburg University. Law*, 11(1), 19–34.
- Basterfield, C., Lilienfeld, S. O., Bowes, S. M., & Costello, T. H. (2020). The Nobel disease: When intelligence fails to protect against irrationality. *Skeptical Inquirer*, 44(3), 32–37.
- Bateson, G. (1972). *Steps to an ecology of mind*. Northvale, NJ: Jason Aronson Inc.
- Bernstein, J. (1978). *Experiencing science*. New York: Basic Books.
- Bernstein, J. (1982). *Science observed: essays out of my mind*. In: American Institute of Physics. New York: Basic Books.
- Bernstein, J. (1990). *Uomini e machine intelligenti*. Milano: Adelphi. (In Ital.).
- Buocz, T. J. (2018). Artificial Intelligence in Court. Legitimacy Problems of AI Assistance in the Judiciary. *Retskraft – Copenhagen Journal of Legal Studies*, 2(1), 41–59.
- Chiodi, G. M. (1979). *La menzogna del potere: la struttura elementare del potere nel sistema politico*. Giuffrè.
- Copeland, B. J. (2004). *The essential turing*. Oxford University Press. <https://doi.org/10.1093/oso/9780198250791.001.0001>
- Crick, F. (1955). *On degenerate templates and the adaptor hypothesis*. Cambridge, England. <https://digirepo.nlm.nih.gov/ext/document/101584582X73/PDF/101584582X73.pdf>
- Crick, F. H. C. (1979). Thinking about the Brain. *Scientific American Magazin*, 241(3), 219–233. <https://doi.org/10.1038/scientificamerican0979-219>
- Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011a). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17), 6889–6892. <https://doi.org/10.1073/pnas.1018033108>
- Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011b). Reply to Weinshall-Margel and Shapard: Extraneous factors in judicial decisions persist. *Proceedings of the National Academy of Sciences*, 108(42), E834–E834. <https://doi.org/10.1073/pnas.1112190108>
- Di Salvo, M. (2024). *Multiple intelligences and specialization. A return to Gardner and the wisdom of rethinking psychology in Hillmanian transparency*. (under publication).
- Dyson, F. J. (1985). *Origins of life*. Cambridge University Press.
- Dyson, G. (2012). *Turing's cathedral: the origins of the digital universe*. Vintage Books.
- European Commission. (2019, April). *Ethical guidelines for trustworthy AI*. <https://ec.europa.eu/newsroom/dae/redirection/document/60419>
- Feroni, G. (2020, December 3). *Algorithms and artificial intelligence. Men make mistakes but this is what makes them higher than the machines*. <https://www.privacy365.eu/en/by-the-italian-data-protection-authority-algorithms-and-artificial-intelligence-men-make-mistakes-but-this-is-what-makes-them-higher-than-the-machines-a-speech-of-ginevra-cerrina/>
- Freeman, J. D. (1970). *The twenty-first century (Speech)*. Vanuxem Lecture. Princeton University.
- Gardner, H. (1983). *Formae Mentis. Essay on the plurality of intelligence*. New York, NY.
- Huxley, A. (1932). *Brave New World*. HarperCollins Publishers.
- Jefferson, G. (1949). The mind of mechanical man. *British Medical Journal*, 1(4616), 1105–1110. <https://doi.org/10.1136/bmj.1.4616.1105>
- McAfee, A., & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future*. WW Norton & Company.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of Mathematical Biophysics*, 5, 115–133. <https://doi.org/10.1007/BF02478259>
- Minsky, M. (1965). *Matter, mind and models*. https://dspace.mit.edu/bitstream/handle/1721.1/6119/AIM-077.pdf?sequence=2&origin=publication_detail
- Mobilio, G. (2020). L'intelligenza artificiale e i rischi di una “disruption” della regolamentazione giuridica. *BioLaw Journal-Rivista di BioDiritto*, 2, 401–424. <https://teseo.unitn.it/biolaw/article/view/1563>
- Morhat, P. M. (2018). *Legal personality of artificial intelligence in the sphere of intellectual property rights: civil legal issues*. (Dr. Sci. (Law). Russian State Academy of Intellectual Property. Moscow. <https://www.dissercat.com/content/pravosubektnost-iskusstvennogo-intellekta-v-sfere-prava-intellektualnoi-sobstvennosti-grazhd?ysclid=lt3zkrhzzk262953893>
- Morozov, E. (2012). *The net delusion: The dark side of Internet freedom*. PublicAffairs.
- Neumann, J. V. (1966). *Theory of self-reproducing automata*. University of Illinois Press.
- Pasquale, F. (2015). *The black box society. The secret algorithms that control money and information*. Harvard University Press.
- Pinker, S. (2022). *Rationality: What it is, why it seems scarce, why it matters*. Penguin Books.
- Schmitt, C. (1927). The Concept of the Political. *Archiv für Sozialwissenschaft und Sozialpolitik*, 58(1), 1–33.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424. <https://doi.org/10.1017/s0140525x00005756>
- Shannon, C. E. (1950). XXII. Programming a computer for playing chess. The London, Edinburgh. *Dublin Philosophical Magazine Journal of Science*, 41(314), 256–275. <https://doi.org/10.1080/14786445008521796>
- Simoncini, A. (2019). L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà. *BioLaw Journal – Rivista di BioDiritto*, 1, 63–89. <https://doi.org/10.15168/2284-4503-352>

- Tegmark, M. (2018). *Life 3.0: Being human in the age of artificial intelligence*. Vintage.
- Tononi, G. (2012). *Phi: A Voyage from the Brain to the Soul*. Pantheon.
- Turing, A. M. (1946). *Proposed electronic calculator. Report*: https://www.alanturing.net/turing_archive/archive/p/p01/p01.php
- Turing, A. M. (1948). *Intelligent machinery: a report*. https://www.alanturing.net/turing_archive/archive/1/132/132.php
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, LIX(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- von Neumann, J. (1958). *The Computer and the Brain*. New Haven: Yale University Press.
- Von Neumann, J. (2014). *Computer e cervello*. Il Saggiatore.
- Weinshall-Margel, K., & Shapard, J. (2011). Overlooked factors in the analysis of parole decisions. *Proceedings of the National Academy of Sciences*, 108(42), E833–E833. <https://doi.org/10.1073/pnas.1110910108>
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Harvard Business School.

Конфликт интересов / Conflict of Interest

Автором не заявлен / No conflict of interest is declared by the author

История статьи / Article history

Дата поступления / Received 12.01.2024

Дата одобрения после рецензирования / Date of approval after reviewing 15.02.2024

Дата принятия в печать / Accepted 20.02.2024