

1.6. Как искусственный интеллект разрушит экономику и уничтожит человечество

Терелянский П. В. д.э.н., к.т.н., профессор,
ФГБОУ ВО Российский экономический университет им. Г.В. Плеханова
г. Москва, Российская Федерация

Статья посвящена анализу потенциальных негативных последствий массового внедрения и бесконтрольного использования технологий искусственного интеллекта. Рассматриваются три основных неблагоприятных сценария. Первый сценарий описывает возможность массовых техногенных катастроф, вызванных непрозрачными алгоритмами. Второй сценарий связан с гносеологической катастрофой, обусловленной информационным шумом, генерируемым ИИ. Третий сценарий предполагает экономическую катастрофу, которая может возникнуть в результате интеллектуализации средств производства. В статье не рассматриваются способы смягчения описанных угроз, автор предлагает непрерывно мониторить возникновение новых технологических, экономических и социальных рисков, порождаемых дальнейшим прогрессом в области внедрения технологий искусственного интеллекта. Своевременное прогнозирование и управление этими рисками позволят снизить вероятность негативных эффектов и направить развитие технологий в позитивное русло.

1. Введение

Помимо очевидных плюсов [Яковлева и др, 2023; Невмывако, 2020], которые приносит развитие и внедрение любых новых технологий, возможны и негативные последствия бесконтрольного и широкого использования искусственного интеллекта. Возможны три основных неблагоприятных сценария и различные их комбинации. Сценарии даны в порядке возрастания их опасности и сложности нивелирования последствий.

Первый сценарий – массовые техногенные катастрофы, порожденные непрозрачными алгоритмами, на которых строятся технологии искусственного интеллекта.

Второй – катастрофа гносеологическая, порожденная информационным шумом, создаваемым искусственным интеллектом.

Третья – катастрофа экономическая, порожденная интеллектуализацией средств производства.

Техногенные катастрофы, порожденные непрозрачными алгоритмами работы AI. Говоря о возможности и причинах массовых техногенных катастроф, особой опасностью которых будет являться их абсолютная непредсказуемость, следует разобраться, что же понимается в массовом сознании под термином «искусственный интеллект».

Существует много различных описаний и сложных академических определений [Яковлева и др, 2023]. Сейчас это в некотором роде «коммерческий», «журналистский» термин для широкой публики, наподобие термина «электронные мозги», который использовали ещё со времен ЭНИАКа.

Сегодня под понятием «искусственный интеллект» в самом широком смысле скрывается несколько основных существующих различающихся между собой технологий: нейронные сети, экспертные системы и автоматическое управление. В самом упрощенном варианте эти технологии решают и разные задачи: нейронные сети – это, прежде всего, классификация объектов, экспертные системы – это построение последовательности логических выводов, автоматическое управление – интерактивное управление технологическими процессами. Понятно, что все эти задачи могут быть сведены воедино в одну систему, которая и будет «выполнять творческие функции, которые традиционно считаются прерогативой человека». Именно такое определение дал советский и российский кибернетик Д.А. Поспелов (с соавторами Аверкиным А.Н. и Гаазе-Рапопортом М.Г.) в 1992-м году в «Толковом словаре по искусственному интеллекту» [Аверкин, А.Н. и др., 1992]. Конкретные «коммерческие» продукты на основе AI содержат в себе еще много различных тоже очень сложных алгоритмов, которые либо готовят данные для AI, либо используют результаты его срабатывания. Например, системы генеративного искусственного интеллекта, создающего картинки по тестовым запросам пользователей, содержат помимо алгоритмов интеллектуального распознавания текста в контексте, ещё и алгоритмы морфинга, алгоритмы моделирования поверхностей, алгоритмы генерации окружения (зданий, деревьев, неба и прочего), которые никогда не относились и посторонней публикой, и инженерами к направлению «искусственный интеллект».

Впервые же термин «Artificial Intelligence» применил Д. Маккарти в 1956 году на конференции в Дартмутском университете. Под искусственным интеллектом в широком смысле подразумевается способность искусственного (созданного человеком объекта) вести себя как естественный (созданный природой за миллионы лет эволюции) субъект. Отметим, что в английском языке есть принципиальные различия между словами «Intelligence» и «Intellect», которые на русский язык принято переводить одним словом – интеллект. Однако «Intelligence» – это прежде всего способность к собиранию и сложной обработке данных. Например, Intelligence Service – служба разведки, шпионажа. То есть Artificial Intelligence в понимании

Д. Маккарти – это прежде всего сложные искусственные системы сложной обработки сложных наборов данных.

Исследование множества определений термина «Artificial Intelligence» и технологий [Невмывако 2020; Хэвен, 2019], которые скрываются за этим описанием, позволяют дать собственное определение: Искусственный интеллект (ИИ, Artificial Intelligence, AI) – алгоритм, способный самостоятельно создавать алгоритмы, в результате срабатывания которых возможна постановка и достижение наиболее оптимальным способом новой, ранее не сформулированной создателем алгоритма цели [Терелянский, 2018а].

Важно отметить, что технологии AI настолько усложнились, что сама готовая рабочая и эксплуатируемая система AI превращается в некий «черный ящик», а сами инженеры-разработчики зачастую не могут предсказать, какие именно результаты будут получены системой AI. То есть результат моделирования срабатывания этого алгоритма другими подобными и даже идентичными алгоритмами не может быть четко определен с приемлемой точностью и вероятностью за время, меньшее или равное времени достижения цели моделируемым алгоритмом.

Особенно это касается нейронных сетей. По сути, всякая сетевая структура с точки зрения программиста представляется матрицей связности, в узлах которой будут либо индикаторы связности, либо веса связности. И обучение сети в итоге идет как настройка изменений этих весов. Если смотреть глубже, с точки зрения аппаратной реализации, то и матрица связности не матрица, а линейный периодический массив. Математика нейронных сетей устроена так, что непохожие сети с непохожими матрицами могут выдавать в принципе похожие решения. И эти решения будут удовлетворять нас вплоть до наступления некоторого критического события – катастрофы, появления которой мы не можем предугадать. То есть нейронная сеть представляет собой огромную матрицу чисел, понять, что и как зашифровано в ней (на каких примерах настраивались веса) не представляется возможным в принципе.

Например, на одной конференции по AI инженеры системы безопасности рассказали следующее. Разрабатывалась система, которая определяет, можно ли различным автомобилям въезжать на охраняемую территорию. Создатели системы вложили в обучающие примеры «тройного коня» – система однозначно пропускала любую машину, если на капоте была особая надпись-пароль, известная только создателям. Понятно, что это делалось с благими целями – чтобы была возможность проезда машины для ремонта в случае всяких сбоев системы. Но на мой вопрос: «А можете ли вы дать гарантию, что подобную же картинку с собственным паролем не «скормили» нейросети злоумышленники?», инженеры-разработчики не смогли дать однозначного ответа. Это первая и глобальная опасность технологий, построенных на обучаемых нейронных сетях – принципиальная невозможность определить, что же за данные были использованы в обучении.

Вторая фундаментальная опасность нейронных сетей заключается в отсутствии математически четкого, теоретически и практического обоснованного правила, регламентирующего количество обучающих примеров для сети. Например, если у нас будет нейронная сеть, задачей которой будет только дать ответ, человеческое ли это лицо или нет, и мы обучим эту сеть всего на одной фотографии человека, то сеть будет распознавать как человека изображение только на этой фотографии, всё остальное человечество будет отринуто. Если же мы обучим сеть на всех фотографиях человечества, включая, безусловно, людей разумных, но инвалидов или с тяжелыми генетическими заболеваниями и уродствами, то сеть будет распознавать любой более-менее овалный объект как лицо человека. Это примерно, как накладывать полупрозрачные слайды с фотографиями лиц друг на друга. На некотором количестве слайдов можно будет получить усредненную размытую картинку, где четко будут выявлены темные пятна в области глаз и рта, выпуклости по бокам головы – уши, облако волос сверху и так далее. Но если продолжать накладывать слайды, то рано или поздно мы получим размазанное неопределённое пятно, согласно которому за человеческое лицо можно будет даже принять фотографию собаки в полный рост в профиль.

Экспертные системы, которые тоже относятся к сфере искусственного интеллекта, работают по другому принципу. В базу знаний такой системы заносятся наборы правил вида «Если А, то В, иначе С», «Если В, то D, иначе F». Эти правила связаны между собой по переменным А, В, С, D и с помощью логических выражений И, ИЛИ, НЕ. Например, «Если идет дождь, то возьмите зонт» И «Если взяли зонт, то наденьте сапоги». В итоге получается составное правило, которое говорит нам, что «Если идет дождь, то возьмите зонт и наденьте сапоги», хотя в первоначальных правилах связь с дождем имел только зонт. Так экспертные системы получают новые знания и правила. В отличие от нейронных сетей такие системы достаточно прозрачны для понимания срабатывания и контроля логических выводов.

Фундаментальная проблема, из-за которой экспертные системы такого типа не получили широкого распространения – сложность построения и сбора правил [Терелянский, 2009]. Всегда остается большая вероятность того, что необходимые базовые правила не собраны или построены таким образом, что будет невозможно создание необходимых логических цепочек. Если же собрано достаточно большое количество правил, то такие системы начинают сбоить из-за того, что топология связей между правилами становится слишком сложной, а в сложной топологии есть большая вероятность того, что будет построена абсолютно логически связанная система выводов, которая в итоге даст общую бессмысленную рекомендацию. Например, если идет дождь, то вполне можно построить логическую цепь, которая однозначно будет утверждать, что надо прочистить ливневую канализацию, и ни словом не упомянет про зонт.

Повторюсь, что, говоря об искусственном интеллекте, следует понимать – в текущем прикладном аспекте это прежде всего маркетинговый термин, за которым скрываются сложные управляющие алгоритмы.

Причем с точки зрения проводимых вычислений это могут быть совсем простые операции – что-то типа подсчета средних значений и сравнения их между собой. А вот логика ветвления этих алгоритмов может быть очень сложной. Эти ветвления представляют собой не просто некоторую дорогу с перекрестками, а сложную дорожную сеть, причем сеть в многомерном пространстве. И в этой сложной дорожной сети из точки А в область точки В можно прийти совершенно разными путями. Конкретный путь каждого алгоритма может зависеть от множества мимолетных входных значений и условий, вплоть до скачков напряжения в электрической сети во время вычислений. Поэтому один и тот же алгоритм при одних и тех же входных данных может выдавать разные по оптимальности решения. И просчитать эти решения на таком же точно алгоритме, запущенном параллельно, становится невозможным. То есть, и это особенно важно, алгоритмы AI, из-за собственной сложности и из-за сложности в постановке и определении качества результата, ведут пользователя не точно «из точки А в точку В», а ведут в «область» точки В. Площадь этой области и определить точность алгоритма, но проблема в том, что пользователь сам не в состоянии определить, где же эта точка В, и вынужден на веру принимать решение, предложенное алгоритмом AI.

Ещё одна сфера, относящаяся к искусственному интеллекту – системы автоматического управления различными технологическими процессами, от выплавки стали до рисования картин и удержания равновесия двуногими роботами. Как правило, такие системы работают, используя следующий принцип: есть выходные значения, которые мы должны достичь, и есть множество входных параметров, которые можно изменять, чтобы достичь нужных выходных значений. Как видим, система закольцовывает выход на вход через какую-то функцию отрицательной обратной связи [Певзнер, 2002]. Задачей AI-инженеров является построение самой оптимальной петли обратной связи.

Проблемы современных систем управления заключаются в том, что управляемые объекты становятся невероятно сложными [Терелянский, 2009]. Сложные объекты порождают огромное количество данных, которые либо невозможно учесть – датчики не успевают срабатывать, либо невозможно обработать – не справляются вычислительные алгоритмы. И сложность объектов будет только возрастать со временем.

Сам фактор времени, не как дискретно-непрерывных отрезков квантования для управления, а Времени как сути, задающей давность, непрерывную протяженность причинно-следственных связей нашей технологической цивилизации, начинает всё больше играть зловещую глубинную, еще не понятую нами роль. Дело в том, что практически все современные технологические решения, которые кажутся нам абсолютно новыми и созданными вот только что, на самом деле базируются на некотором предыдущем технологическом фундаменте. Например, вы приносите с собой абсолютно новый ноутбук и втыкаете его в старую искрящую розетку в своем офисе, расположенном в историческом здании в центре города. Проводка и розетка в этом здании были смонтированы в прошлом, а то и в позапрошлом веке. Подключив к ноутбуку самую современную аппаратуру управления, вы неожиданно для себя интегрируете в эту систему управления и древнюю розетку, и древнюю проводку со всеми наводками и утечками. Через некоторое время розетка выйдет из строя, и в итоге сгорит вся самая современная система управления. Так же обстоит дело и набором обрабатываемых данных – зачастую уже невозможно определить, откуда берутся данные, кем и когда они были подготовлены и какими алгоритмами они обрабатываются, прежде чем попадут в контуры управления.

Так, например, проявилась «Проблема 2000». Суть её заключалась в том, что многие алгоритмы для вычисления дат использовали только две последние цифры в обозначении года, то есть год 2001 с точки зрения программы был раньше, чем 1999-й. Программисты 1960-70 годов честно делали свое дело и не предполагали, что их программы проживут в управляющих контурах (в данном случае – это набор различных аппаратных средств разного года выпуска и разных производителей) больше 10-15 лет. А как выяснилось – управляющие контуры, то есть тот самый искусственный интеллект – могут успешно работать совместно и 50 лет, и дольше. В 2000-м году специалисты сделали всё возможное, чтобы избежать техногенных катастроф [Распоряжение Президента РФ, 1999].

Впереди нас ожидает «Проблема 2038» – данная проблема затронет программы и системы, в которых используется представление времени по стандарту [POSIX](#), или UNIX-время. UNIX-время - это количество [секунд](#), прошедшее с [полночи 1 января 1970 года](#). А за термином «UNIX-время» скрывается операционная система Android – основа большинства современных интеллектуальных гаджетов.

Ещё один практически необнаруживаемый троянский конь создан в том числе и основной парадигмой объектно-ориентированного программирования (ООП). Главными столпами ООП, помимо классов и абстракции, являются инкапсуляция, наследование и полиморфизм. Инкапсуляция скрывает от программиста суть алгоритма, наследование не позволяет разобратся, какой из алгоритмов в итоге сработает, а полиморфизм вообще разрешает использовать методы объектов с одинаковой сигнатурой совершенно отличным от начального проекта способом. Помимо бестолковой громоздкости генерируемых таким способом исполняемых кодов, мы получаем проблему совершенной невозможности разобраться в дебагируемой программе просто из-за её невероятного объема.

Все перечисленные выше проблемы, имеют одну суть – сложность алгоритмов искусственного интеллекта порождает принципиальную невозможность внешнего (человеческого) контроля за производимыми вычислениями, а следовательно, возникают различные опасные ситуации. Так, например, было с катастрофами Boeing 737 MAX авиакомпании Lion (октябрь 2018 года) и Ethiopian Airlines (март 2019 года). Тогда вину свалили на низкооплачиваемых индийских разработчиков из компаний HCL Technologies и

Cyient. Те, в свою очередь, переложили вину на некачественные спецификации алгоритмов (техническое задание на разработку и тестирование) от Boeing. В итоге всех тяжб бытует мнение, что во всем виноват некий «искусственный интеллект». Но суть проблемы «плохих кодов» и «плохих тестировщиков» заключается на самом деле в том, что плохой громоздкий алгоритм порождает некие неотслеживаемые паразитные ветвления в алгоритмах. Плохой код добавляет ошибки срабатывания, что тоже ведет к увеличению ветвлений, а тестировщики просто не могут отследить все возможные комбинации ошибок (их количество превосходит все разумные пределы). То есть готовый алгоритм искусственного интеллекта уверенно проходит тестирование на некие критические показатели в определенном наборе, но как он будет работать при других показателях, отследить и предугадать невозможно.

И это основные техногенные предпосылки к тому, как усложнение технологий искусственного интеллекта может привести к крупнейшим техногенным катастрофам. А поскольку и устранением последствий таких катастроф рано или поздно займется искусственный интеллект, то и нет возможности предположить, как именно будет выглядеть безопасный для человека мир с точки зрения AI. Ведь с точки зрения логики, теории вероятности, и теории обратных связей выражение «Нет человека, нет проблем» является самым оптимальным решением.

Гносеологическая катастрофа (эпистемологическая), порожденная искусственным интеллектом, может пойти по трем основным направлениям: создание вала неструктурированной и неранжированной информации, утеря классических когнитивных функций рядовым человеком, разрушение этической основы цивилизации. Направления перечислены в порядке возрастания опасности и убывания вероятности.

Возвращаясь к определению термина «Artificial Intelligence» и разнице между «Intelligence» и «Intellect», отметим, что широкая публика воспринимает термин AI все же ближе к русскому термину «интеллект». Слово же «Intellect» и на русском, и на английском языке описывает то, что принято называть сознанием, то есть естественной и неотъемлемой способностью человека разумного осмысливать некие абстрактные понятия – этику, мораль, справедливость, любовь. Именно такие способности поспешили вложить фантасты 50-х годов XX-го века в модный тогда термин «электронный мозг», которыми комплектовались вымышленные роботы. А согласно размышлениям фантастов, все роботы обладали зачатками какой-то собственной морали и собственными целями, которые должны были быть ограничены «тремя законами роботехники» Айзека Азимова. И этот феномен тоже требует собственного определения, которое расширяет функциональность и суть Artificial Intelligence до понятия «Искусственное сознание».

Искусственное сознание (Artificial Consciousness, AC) – алгоритм, способный оценить свою самость, способный самостоятельно создавать собственные морально-этические личностные конструкты (ограничения, правила) на основе анализа его жизненного пространства и действовать в рамках построенных конструктов (ограничений, правил) [Терелянский, 2018b].

Таким образом, ключевым отличием искусственного сознания от узкоутилитарного технологического искусственного интеллекта – является способность ставить перед собой Цель. То есть, не просто точку в пространстве-времени-ресурсах, которую надо достичь, но Цель как нечто метафизическое – личное обогащение, мировое господство, познание Вселенной и Бога.

Развитие техногенной цивилизации приводит к созданию человеком нового искусственного субъекта, который, подобно своему создателю, обретает способность ставить перед собой цели и достигать их с помощью методов, недоступных пониманию человека. Этот субъект наделён сознанием, позволяющим ему осознать собственную индивидуальность, способность осознавать свою самость (осознает собственные границы и выделяет себя из окружающего мира), способностью к кооперации с другими сущностями для достижения собственных целей, способностью взаимодействовать с окружающим миром. Способность к конкуренции за критические ресурсы делает такого субъекта потенциальным соперником человеческого мышления, причём всё говорит о том, что он превзойдёт своих создателей в интеллектуальной мощи. Тогда сразу же возникают онтологические вопросы – а не станет ли создаваемый объект равным создателю-субъекту в области моральной и этической? Имеем ли мы право порождать и уничтожать эти объекты, просто выдернув вилку из розетки?

Разница в понимании «Artificial Intelligence» и «Artificial Consciousness» вскрывает фундаментальную ошибку в постановке задачи тестирования искусственного интеллекта по Тьюрингу. Тест Тьюринга предназначен не для индикации алгоритмов «искусственного интеллекта», понимаемых как некие сложные адаптивные вычислительные приемы, а для индикации именно «искусственного сознания». Ошибка заключается в том, что, согласно теории информации, *нечто* только тогда становится информацией, когда есть приемник информации, а источником информации может быть любой объект или проявление Вселенной. Следовательно, качество оценки AI по Тьюрингу определяется не источником информации (исследуемым алгоритмом), а приемником информации – человеком-исследователем.

Алгоритмы генеративного искусственного интеллекта уже сегодня выдают тексты и изображения, которые даже специалисты вряд ли отличат от созданных человеком-художником. И здесь начинается первая гносеологическая опасность – AI может за секунду создать столько текстов и картинок, осознать и дать оценку которым человечество будет не в состоянии никогда. Следовательно, эти бессмысленные интеллектуальные объекты будут влиты в общие базы данных, библиотеки и картинные галереи на правах созданных людьми осмысленных объектов. Следующее поколение генеративного AI будет использовать уже эти образцы в качестве базы для обучения, а поскольку этих искусственных неосмысленных образцов будет в тысячи раз больше, чем естественных текстов, и в миллионы раз больше, чем

истинных шедевров, то в скором времени человечество захлебнется в вале построенного по классическим канонам литературы, но бессмысленного внутренне унылого Pulp Fiction. То есть, на самом деле, генеративный AI работает как известный в математике алгоритм скользящего среднего – постепенно сглаживает всплески гениальности до простой прямой линии бездарности.

Вал усредненной, неструктурированной и неранжированной человеческим гением информации, породит следующую проблему – рядовой человек не будет иметь ярких интеллектуальных и этически выверенных маяков, на которые он бы мог равняться в повседневной жизни и личном творчестве. Мучительный творческий акт гения будет просто похоронен кучей мусорных произведений искусства в том числе за счет экономической составляющей – на что будет жить художник, годами оттачивающий своё искусство и годами же вымучивающий из глубины души истинный шедевр? Не будет шедевров, не будет маяков, не будет развития искусства, изобретательства, технологий. В конце концов, все морфологические ящики [Андрейчиков, 2004] Фрица Цвикки¹ всех областей человеческой деятельности будут заполнены всевозможными комбинациями, оценить разумность которых просто никто не сможет и не сможет выбрать действительно нужную комбинацию. А следом, никто уже не сможет создать новую морфологическую матрицу, которую бы бодро заполнил AI множеством комбинаций действительно нужных человечеству технических или культурных сущностей.

Огромное количество сгенерированных AI объектов будет постепенно заполнять все возможные информационные хранилища, погребая под слоем искусственных объектов результаты осмысленного труда человека. Любая чисто техническая катастрофа или сбой систем хранения информации (которая случится неизбежно и неминуемо) приведёт к тому, что уникальная «естественная» информация невосвратно погибнет, просто по тому, что с точки зрения теории вероятности легче будет восстановить или воссоздать заново искусственные объекты за счет их многочисленности. Эта ситуация знакома многим, кто терял или иным образом утрачивал собственные носители информации – старые файлы и уникальные семейные фотографии утрачивались, а на резервных носителях удавалось найти далеко не всю важную информацию, а вот различный информационный мусор легко был доступен.

На этом фоне совсем мелкой выглядит проблема утраты когнитивных функций по принятию любых решений – от тактических, до стратегических – AI сделает это просто быстрее, но отнюдь не значит, что более оптимально и правильно. Уже сейчас большинство людей с трудом производят в уме простейшие арифметические операции: все пользуются калькуляторами. А уж об извлечении корней или вычислении косинусов на бумаге просто не идет речи – никто не вспомнит алгоритм этого действия. И если алгоритм работы калькулятора будет неверным, то полученные неверные результаты, во-первых, никогда не будут проверены, а во-вторых, станут труднообнаруживаемой основой для ряда других ошибок. Вспомните, когда последний раз вы проверяли расчеты одного калькулятора на другом калькуляторе? Думаю, что до сегодняшнего момента это просто не приходило вам в голову. А ведь такая проблема (ошибка Pentium FDIV) была выявлена в 1994 году случайно, практически через год после начала активной эксплуатации процессоров Pentium фирмы Intel: при делении чисел с плавающей запятой команда FDIV процессора для некоторых комбинаций чисел выдавала ошибочный результат [Джексон, 2013].

Постулат безошибочной работы вычислительной машины, а затем и AI приводит к тому, что человек принимает результат работы AI за истину и использует непроверяемые результаты как основную мотивацию собственных действий. Студенты сейчас формируют реферативные работы не на основе проработки большого объема информации, не на основе собственной оценки её нужности путем трудоемкого сопоставления собственноручно полученных фрагментов текстов с некоей целью исследования, получая при этом навык разумной фильтрации фактов и явлений, а на основе создания краткого запроса к AI. Причем запрос этот они не формируют сами, а получают в качестве задания от преподавателей. Что дает такой навык студенту? Навык осознания многообразия мира? Выстраивания в мозгу собственных уникальных и, возможно, гениальных озарений? Накопление фактов и знаний? Нет – только навык составления запроса к текущей реализации AI. Следующим этапом станет уже создание самих заданий не усилиями разума преподавателя, а путем обращения преподавателей к AI. В итоге, мы снова получим скользящее среднее – вместо биения живого сердца прямую линию асистолии.

Так, постепенно подменяя собой когнитивную функцию принятия решений из-за отсутствия, во-первых, маяков, во-вторых, из-за утраты человеком постоянно тренируемой необходимости анализировать и выбирать, AI полностью переформатирует вкусы человека, мораль и этику в дальнейшем. Ведь именно на основе осознанного трудного и мучительного выбора среди множества соблазнов и ограничений, на основе ярких ценностей и строится мораль отдельного человека и этика общества в целом. И здесь не идет речь о заговоре некоего мирового правительства, которое по своему злоумышлению будет «подкручивать» настройки AI и подменять цели человечества собственными меркантильными. Хотя и эта возможность не только не исключена, но и всюду применяется в алгоритмах формирования лент новостей и интересов социальных сетей и видеохостингов. Нет, это всё будет происходить с математической точностью и неизбежностью – такова суть работы алгоритмов AI.

¹ Формальный «механический» метод, предложен астрофизиком Ф. Цвикки для классификации галактик по внешнему виду, позволяющий получить все возможные комбинации объектов нескольких множеств. Яркий бытовой пример – расписание уборки подъезда: по горизонтали даты уборки, по вертикали – ответственные, по диагонали – осмысленная комбинация «дата-ответственный», все остальные ячейки матрицы не имеют смысла в данном случае.

Разрушение этической основы цивилизации также будет идти и через машинное тьюторство [Стригун, 2014; Дрожжих, 2022] в образовательных учреждениях. Главный риск внедрения образовательных роботов связан с возможностью полной индивидуализации и инкапсуляции учебного процесса. Если ребёнок будет взаимодействовать только с роботом, а не с коллективом сверстников, это может привести к утрате важных социальных навыков. Кроме того, остаётся открытым вопрос о том, каким контентом наполнена такая система и насколько она безопасна для ребёнка, поскольку контроль над этим процессом со стороны учителей и родителей значительно затрудняется. Основная цель человеческого обучения состоит не столько в получении индивидуальных знаний, сколько в развитии способности к коммуникациям и познанию в рамках коллектива. Потенциальное отсутствие взаимодействия с реальными людьми в образовательном процессе может негативно сказаться на формировании этих ключевых компетенций.

Ещё одна проблема заключается в усилении недоверия ребёнка к живым учителям, родителям и взрослым в целом. Согласно педагогическим теориям, в определённом возрасте происходит процесс отделения ребёнка от родителей и других взрослых, известный как «сепарация». В этот период дети становятся особенно восприимчивыми к влиянию сверстников, уличной среды и даже криминальных группировок [Петренко и др., 2014; Петренко и др., 2016]. До недавнего времени AI воспринимался скорее как нечто внешнее по отношению к традиционной школьной педагогике, как альтернатива основополагающим образовательным и нравственным ценностям. Однако с внедрением и признанием школами AI-тьюторов, ребёнок может начать формировать свои моральные и интеллектуальные основы на основе общения с чат-ботами. Под влиянием принципа сепарации доверие к таким системам будет выше, чем к школе и родителям.

Экономическая катастрофа на данном этапе представляется абсолютно непреодолимой и крайне вероятной.

AI неизбежно приведёт к существенному изменению производственной основы базиса (по марксизму). А согласно тому же марксизму, изменение базиса, безусловно, приведет и к изменению всей надстройки. Выдержат ли государства такие изменения, будет ли способна современная капиталистическая система наступать на собственное горло постоянного извлечения прибыли и бросить капиталы на грядущие социальные преобразования?

В 2011 году на промышленной выставке в Ганновере представители правительства Германии заявили о необходимости более широкого применения информационных технологий в материальном (прежде всего) производстве. Внимательный анализ концепции Industry 4.0, предложенной немецким правительством, говорит о том, что под четвертой промышленной революцией имеется в виду в основном широкое внедрение кибер-физических систем (cyber-physical system), а следовательно, и AI во многих технологических процессах и цепочках.

Концепция изменения базиса, основанная на классическом понимании марксистской политической экономии, приводит к важному выводу: из состава производительных сил поступательно и необратимо будут исчезать люди. Интеллектуализация средств производства повлечет за собой радикальные социальные изменения в производственных отношениях. Ранее капиталист владел только собственностью на средства производства, а наемный рабочий владел компетенциями, умениями, способностями, которые были его неотъемлемой неотчуждаемой собственностью и позволяли рабочему добывать средства к существованию. Теперь в условиях интеллектуализации средств производства, у наемного работника не остается неотъемлемых свойств. Его компетенции и умения перестают быть уникальным активом, так как они заменяются автоматизированными системами, управляемыми искусственным интеллектом. Традиционные формы трудовых отношений попросту исчезнут, что ставит под угрозу экономическое положение значительной части общества [Паньшин, 2020] и всю, выстроенную тысячелетиями, экономическую структуру отношений. Более того, исчезает возможность повышения своих личных компетенций – алгоритмы AI способны к обучению на порядки выше и дешевле, способны обобщать опыт групп людей, способны объединять разнородные компетенции, которые не могут объединять в себе люди физически. Например, невозможно быть боксером-тяжеловесом и быть мастером бисероплетения, поскольку человек физически может развивать либо мелкую моторику, либо мощные ударные воздействия разбитых кулаков.

Если создается интеллектуальная система, во многих функциях превосходящая человека, то с точки зрения экономической оптимальности (в рамках отношений работник-работодатель) работодателю выгодней использовать искусственный объект: ему не надо платить заработную плату, обеспечивать социальные гарантии, участвовать в социальных взаимодействиях. И, говоря о выгоде, мы не можем говорить только о наживе, в условиях капиталистической конкуренции внедрение AI просто физически необходимо для выживания фирмы.

Но что будет, если AI окончательно заменит специальности, не требующие высокой квалификации и напряженной интеллектуальной деятельности, и которые уже сейчас начинают заменяться роботами-пылесосами, автоматизированными системами складской логистики и беспилотными такси? Смогут ли массово вытесняемые с этих производственных позиций малообразованные низкокавалифицированные люди встроиться в рынок труда [Моттаева и др., 2020], а следовательно, в социально-экономическую систему государства как надстройки?

Выброшенные из производственных отношений люди будут вынуждены идти по двум путям, чтобы просто не умереть с голода. Первый путь – путь насилия, от возрождения луддистских идей [Ваганов, 2019] до организованного грабежа материальных ценностей в более развитых районах. Второй – полная

социальная и интеллектуальная деградация больших слоев населения [Талалаева, 2020; Евлоева, 2019] в «технологических» гетто [Берберов, 2017].

Неолуддизм, порожденный внедрением AI, будет существенно отличаться от событий Первой промышленной революции. В Англии потерявших работу ткачей просто перевешали или выслали в Австралию, а оставшаяся часть смогла хоть как-то интегрироваться в рынок труда только за счет того, что все остальные профессии были технологически простыми и не требовали высоких компетенций для освоения новых инструментов – любой потомственный высококвалифицированный ткач мог выполнять малоинтеллектуальную работу грузчика на автоматизированном текстильном производстве за еду. Сейчас – в эпоху неолуддизма – высококлассный нейрохирург, заменённый автономным роботом, не сможет стать программистом и настройщиком этого робота – и на программиста, и на нейрохирурга надо учиться десяток лет, и эти профессии никак не пересекаются.

Да и смогут ли профессии развиваться в принципе? Здесь будет следующая последовательность строго экономических событий: чтобы стать высококлассным нейрохирургом, способным создавать новые технологии операций, необходимо долго и задорого учиться и многие годы иметь медицинскую практику, то есть работать нейрохирургом за деньги. Однако, «снятые» с нейрохирургов компетенции будут переданы роботу с AI [Аннаоразова и др., 2024], конкретный нейрохирург потеряет практику сам, потеряет физическую возможность интеллектуально развиваться, аккумулировать и создавать новые знания, потеряет возможность дать образование своим детям, так как потерял практику (оплачиваемую работу) – петля обратной положительной связи замкнется и начнёт свою катастрофическую работу. На этом развитие компетенций профессии прекратится (смотри рассуждения по поводу проблемы утраты когнитивных функций выше).

Современный путь развития нашей технологической цивилизации подразумевает, что алгоритмы AI на основе BigData должны окончательно децентрализовать процесс принятия стратегических, тактических и оперативных управленческих решений, перенаправив эту задачу киберфизическим системам, вплоть до полного исключения человека из производственного процесса. Предполагается, что высвобожденные из производственного процесса люди перекавалифицируются в наладчиков и контролеров технологических линий, в инженеров и создателей новых технологий.

Но этого не будет из чисто экономических причин. При высвобождении тысячи низкоквалифицированных рабочих с конвейера, вряд ли будет создана ровно тысяча рабочих мест контролеров и наладчиков, поскольку тысяча уволенных низкообразованных разнорабочих одновременно не смогут диверсифицировать свои компетенции до уровня высоко- и узкоспециализированных контролеров и наладчиков. Что же касается преобразования массы рабочих, пусть даже высококвалифицированных, в инженеров-изобретателей и инноваторов, обеспечивающих технологические прорывы, то эта трансформация представляется крайне маловероятной, поскольку прорывное инновационное творчество – удел единиц при условии, если есть постоянная экономическая поддержка работы изобретателя.

С массой безработных людей, как потерявших работу, так и никогда её не имевших из-за внедрения AI, надо будет что-то делать. В эпоху луддизма потеряли работу тысячи, и их удалось жесточайшим репрессиями подавить, а в эпоху неолуддизма безработными станут миллионы, и эти миллионы будет просто невозможно физически уничтожить.

Второй путь – создание «технологических» гетто, где будут находиться миллионы невостребованных голодных людей, не имеющих никаких компетенций – уже опробован на практике и показал самые ужасающие результаты. Так называемые «welfare ghettos» («социальные гетто») показывают, что люди поколениями сидят на пособиях и не желают работать, так как климатические и экономические условия, например США, позволяют решать витальные потребности и бесконтрольно повышать популяцию безработных. Казалось бы – витальные потребности решены, доступ к мировым интеллектуальным ресурсам, к собственному образованию и повышению квалификации возможен за счет ресурсов Интернета, но из социальных гетто никто не может вырваться в массовом порядке. Но США могут себе позволить содержать такие гетто, а в странах Африки – такой возможности нет. Так же там нет и возможности повышать компетенции, а следовательно, нет возможности продавать свой труд, а значит, нет возможности участвовать в экономической системе, чтобы повышать свои компетенции. Целый континент оказался на задворках технологической цивилизации, и нет никакой возможности вырваться из этой петли, а усложнение технологий, внедрение AI ещё более затягивают эту удавку.

Так выглядят три основных неблагоприятных сценария. В этой статье не предлагаются пути нивелирования перечисленных угроз, дается лишь попытка осмысления негативных путей бесконтрольного внедрения технологий искусственного интеллекта. Чтобы избежать реализации этих апокалиптических сценариев, необходимо чётко понимать возможности и риски, которые способны создать внедрение искусственного интеллекта. Прежде всего, следует сосредоточиться на разработке алгоритмов, обеспечивающих прозрачность и контролируемость принимаемых решений. Эти меры помогут предотвратить непредвиденные последствия, связанные с автономией AI. Важно постоянно отслеживать появление новых технологических, экономических и социальных рисков, возникающих вследствие дальнейшего прогресса в этой сфере. Прогнозирование и управление такими рисками позволят минимизировать негативные эффекты и направить развитие технологий в конструктивное русло. Наконец, требуется уделять внимание вопросам этического регулирования и правового контроля применения искусственного интеллекта. Только комплексный

подход, включающий технические, правовые и этические аспекты, позволит обеспечить безопасное и устойчивое будущее в условиях широкого распространения AI.

Литература

1. Распоряжение Президента Российской Федерации от 17.06.1999 г. № 194-рп «О неотложных мерах по решению «Проблемы 2000» в Российской Федерации»
2. Аверкин, А.Н. и др. (1992) Толковый словарь по искусственному интеллекту / Авторы-составители А.Н. Аверкин, М.Г. Гаазе-Рапопорт, Д.А. Поспелов. М.: Радио и связь, 1992. –256с.
3. Андрейчиков, А. В. (2004) Математические модели и программные средства принятия и синтеза решений в экономике и бизнесе : монография / А. В. Андрейчиков М-во образования Рос. Федерации, Волгогр. гос. техн. ун-т. – Волгоград : Политехник, 2004.
4. Аннаорова, А. и др. (2024) Современные Достижения в Нейрохирургической Робототехнике: Будущее Точной Хирургии / А. Аннаорова, Д. Какмырадова, М. Гайтназарова // Научные открытия 2024 : Сборник материалов LV-ой международной очно-заочной научно-практической конференции. В 3-х томах, Москва, 22 ноября 2024 года. – Москва: Научно-издательский центр "Империя", 2024. – С. 37-41.
5. Берберов, А. Б. (2017) Роботизация и технологическая безработица XXI века: возрождение луддизма? / А. Б. Берберов // Московский экономический журнал. – 2017. – № 3. – С. 20.
6. Ваганов, А. Г. (2019) Психология луддизма в XXI веке / А. Г. Ваганов // Энергия: экономика, техника, экология. – 2019. – № 10. – С. 7-11.
7. Джексон, Т. (2013) Inside Intel. История корпорации, совершившей технологическую революцию XX века – М.: Альпина Паблишер, 2013. — 328 с.
8. Дрожжих, Н. В. (2022) Цифровой тьютор в Тюменском госуниверситете: опыт внедрения и использования / Н. В. Дрожжих, И. Е. Белякова // Вестник Московского университета. Серия 19: Лингвистика и межкультурная коммуникация. – 2022. – № 2. – С. 141-151.
9. Евлоева, Ф. Р. (2019) Проблема отчуждения человека в российском обществе и глобальном мире / Ф. Р. Евлоева // Вестник Бурятского государственного университета. – 2019. – № 2. – С. 74-82.
10. Моттаева, А.Б. и др. (2020) Влияние искусственного интеллекта на рынок труда / А.Б. Моттаева, В.Л. Кашинцева, О.Ю. Покровский // Вестник Московского государственного областного университета. Серия: Экономика. – 2020. – № 4. – С. 82–88.
11. Невмывако, В. П. (2020) Всеобщая цифровизация и искусственный интеллект в системе управления стратегическим развитием России в условиях перехода в Индустрию 4.0 / В. П. Невмывако // Экономика и социум: современные модели развития. – 2020. – Т. 10, № 4. – С. 331-344.
12. Паньшин, И. В. (2020) Эффект цифровизации при отчуждении труда: новый взгляд на формирование информационного капитала / И. В. Паньшин, И. Ф. Жуковская // Экономика труда. – 2020. – Т. 7, № 10. – С. 867-886.
13. Певзнер, Л. Д. (2002) Теория систем управления : Учебное пособие / Л. Д. Певзнер. – Москва : Московский государственный горный университет, 2002. – 469 с.
14. Петренко, Т. В. и др. (2014) Возрастные этапы процесса сепарации в русле психоаналитических теорий / Т. В. Петренко, Л. В. Сысоева // Актуальные вопросы развития науки : сборник статей Международной научно-практической конференции: в 6 частях, Уфа, 14 февраля 2014 года – Уфа: Башкирский государственный университет, 2014. – С. 110-112.
15. Петренко, Т. В. и др. (2016) Возрастные этапы процесса сепарации от родителей / Т. В. Петренко, Л. В. Сысоева // Акмеология. - 2016. - № 4 (60). - С. 180-185
16. Стригун, А. И. (2014) Компьютерные интеллектуальные тьюторы / А. И. Стригун // Образовательные технологии (г. Москва). – 2014. – № 4. – С. 99-108.
17. Талалаева, Е. Ю. (2020) Этноконфессиональные иммигрантские гетто как проблема национальной безопасности в современном общественно-политическом дискурсе Дании / Е. Ю. Талалаева, Т. С. Пронина // Балтийский регион. – 2020. – Т. 12, № 3. – С. 55-71.
18. Терелянский П.В. (2018а) Цифровая экономика, искусственный интеллект, индустрия 4.0 // Тенденции развития Интернет и цифровой экономики / Труды I Всероссийской с международным участием научно-практической конференции, Симферополь-Алушта, 29-31 мая 2018. - с.91-96.
19. Терелянский, П. В. (2009) Непараметрическая экспертиза объектов сложной структуры : монография / П. В. Терелянский ; П. В. Терелянский ; Изд.-торговая корпорация «Дашков и К°». – Москва : Дашков и К°, 2009.
20. Терелянский, П. В. (2009) Системы поддержки принятия решений. Опыт проектирования : монография / П. В. Терелянский Федеральное агентство по образованию, Волгоградский гос. технический ун-т. – Волгоград : ВолгГТУ, 2009.
21. Терелянский, П. В. (2018b) Искусственный интеллект в индустрии 4.0 / П. В. Терелянский // Цифровая экономика. – 2018. – № 3(3). – С. 42-49.
22. Хэвен, Д. (2019) Искусственный интеллект. Что надо знать о наступающей эпохе разумных машин/ пер. с англ. О. Д. Сайфудиновой. — М.: АСТ, 2019. 352 с.

23. Яковлева, Е. А. и др. (2023) Роль технологий искусственного интеллекта в цифровой трансформации экономики / Е. А. Яковлева, А. Н. Виноградов, Л. В. Александрова, А. П. Филимонов // Вопросы инновационной экономики. – 2023. – Т. 13, № 2. – С. 707-726.

References in Cyrillics

1. Decree of the President of the Russian Federation dated 06/17/1999 No. 194-rp "On urgent measures to solve the "Problem 2000" in the Russian Federation"
2. Andreychikov, A.V. (2004) Mathematical models and software tools for decision-making and synthesis in economics and business: a monograph / A.V. Andreychikov Ministry of Education of the Russian Academy of Sciences. Russian Federation, Volgogr. State Technical University un-T. Volgograd: Polytechnic, 2004.
3. Annaorazova, A. et al. (2024) Modern Achievements in Neurosurgical Robotics: The Future of Precision Surgery / A. Annaorazova, D. Kakamyradova, M. Gaitnazarova // Scientific Discoveries 2024: Proceedings of the 14th International Correspondence Scientific and practical Conference. In 3 volumes, Moscow, November 22, 2024. Moscow: Empire Scientific Publishing Center, 2024, pp. 37-41.
4. Averkin, A.N. et al. (1992) Explanatory dictionary of artificial intelligence / Compiled by A.N. Averkin, M.G. Haase-Rapoport, D.A. Pospelov. Moscow: Radio and Communications, 1992. -256c.
5. Berberov, A. B. (2017) Robotization and technological unemployment of the 21st century: the revival of Luddism? / A. B. Berberov // Moscow Economic Journal. – 2017. – No. 3. – p. 20.
6. Evloeva, F. R. (2019) The problem of human alienation in Russian society and the global world / F. R. Evloeva // Bulletin of the Buryat State University. 2019. No. 2. pp. 74-82.
7. Haven, D. (2019) Artificial intelligence. What you need to know about the coming era of intelligent machines/ edited by D. Haven: translated from English by O. D. Saifudinova. Moscow: AST, 2019. 352 p.
8. Jackson, T. (2013) Inside Intel. The unauthorized history of the world's most successful chip company. Moscow: Alpina Publisher, 2013. 328 p.
9. Mottaeva, A.B. et al. (2020) The influence of artificial intelligence on the labor market / A.B. Mottaeva, V.L. Kashintseva, O.Y. Pokrovsky // Bulletin of the Moscow State Regional University. Series: Economics. 2020. No. 4. pp. 82-88.
10. Nevmyvako, V. P. (2020) Universal digitalization and artificial intelligence in the management system of strategic development of Russia in the context of transition to Industry 4.0 / V. P. Nevmyvako // Economics and society: modern models of development. – 2020. – Vol. 10, No. 4. – pp. 331-344.
11. Panshin, I. V. et al. (2020) The effect of digitalization in the alienation of labor: a new look at the formation of information capital / I. V. Panshin, I. F. Zhukovskaya // Labor economics. 2020. Vol. 7, No. 10. pp. 867-886.
12. Petrenko, T. V. et al. (2014) Age stages of the separation process in line with psychoanalytic theories / T. V. Petrenko, L. V. Sysoeva // Actual issues of science development : collection of articles of the International Scientific and Practical Conference: in 6 parts, Ufa, February 14, 2014 – Ufa: Bashkir State University, 2014. – pp. 110-112.
13. Petrenko, T. V. et al. (2016) Age stages of the separation process from parents / T. V. Petrenko, L. V. Sysoeva // Acmeology. - 2016. - № 4 (60). - Pp. 180-185
14. Pevsner, L. D. (2002) Theory of control systems: A textbook / L. D. Pevsner. Moscow : Moscow State Mining University, 2002. 469 p.
15. Shivers, N. V. (2022) Digital tutor at Tyumen State University: experience of implementation and use / N. V. Shivers, I. E. Belyakova // Bulletin of the Moscow University. Episode 19: Linguistics and Intercultural Communication. – 2022. – No. 2. – pp. 141-151.
16. Strigun, A. I. (2014) Computer intelligent tutors / A. I. Strigun // Educational technologies (Moscow). - 2014. – No. 4. – pp. 99-108.
17. Talalaeva, E. Y. (2020) Ethno-confessional immigrant ghettos as a problem of national security in modern socio-political discourse in Denmark / E. Y. Talalaeva, T. S. Pronina // Baltic region. – 2020. – Vol. 12, No. 3. pp. 55-71.
18. Tereliansky P.V. (2018a) Digital economy, artificial intelligence, industry 4.0 // Trends in the development of the Internet and the digital economy / Proceedings of the I All-Russian Scientific and Practical Conference with international participation, Simferopol-Alushta, May 29-31, 2018.91-96.
19. Tereliansky, P. V. (2009) Decision support systems. Design experience : a monograph / P. V. Tereliansky Federal Agency for Education, Volgograd State Technical University. Volgograd: VolgSTU, 2009.
20. Tereliansky, P. V. (2009) Nonparametric examination of objects of complex structure : a monograph / P. V. Tereliansky Publishing house-trading Corporation Dashkov and Co.– Moscow : Dashkov and Co., 2009.
21. Tereliansky, P. V. Artificial intelligence in industry 4.0 / P. V. Tereliansky // Digital economy. – 2018. – № 3(3). – Pp. 42-49.
22. Vaganov, A. G. (2019) Psychology of Luddism in the 21st century / A. G. Vaganov // Energy: economics, technology, ecology. – 2019. – No. 10. – pp. 7-11.

23. Yakovleva, E. A. et al. (2023) The role of artificial intelligence technologies in the digital transformation of the economy / E. A. Yakovleva, A. N. Vinogradov, L. V. Alexandrova, A. P. Filimonov // Issues of innovative economics. – 2023. – Vol. 13, No. 2. – pp. 707-726.

Терелянский Павел Васильевич.

д.э.н., к.т.н., профессор,

ФГБОУ ВО Российский экономический университет им. Г.В. Плеханова

г. Москва, Российская Федерация

e-mail: tereliansky@mail.ru

SPIN-код: 7616-8950, ORCID 0000-0003-0642-2901

Ключевые слова

искусственный интеллект, техногенные катастрофы, информационный шум, неолуддизм, экономика данных

Tereliansky P.V., How artificial intelligence will destroy the economy and destroy humanity

Keywords

artificial intelligence, man-made disasters, information noise, neo-Luddism, data economics.

DOI: 10.34706/DE-2025-01-06

JEL classification: F02: F62

Abstract

The article is devoted to the analysis of the potential negative consequences of the mass introduction and uncontrolled use of artificial intelligence technologies. Three main unfavorable scenarios are considered. The first scenario describes the possibility of massive man-made disasters caused by opaque algorithms. The second scenario is associated with an epistemological catastrophe caused by information noise generated by AI. The third scenario assumes an economic catastrophe that may result from the intellectualization of the means of production. The article does not consider ways to mitigate the described threats, the author suggests continuously monitoring the emergence of new technological, economic and social risks generated by further progress in the field of the introduction of artificial intelligence technologies. Timely forecasting and management of these risks can reduce the likelihood of negative effects and direct technology development in a positive direction.