

РАЗВИТИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ГЛОБАЛЬНЫЕ ТРЕНДЫ И НОВЫЕ ВЫЗОВЫ

А. Я. Сарна

*Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, alsar.05@mail.ru*

В статье рассматриваются особенности искусственного интеллекта в его сравнении с «естественным» на основе гипотетической возможности прохождения теста А. Тьюринга не только человеком, но и «умным» техническим устройством – системой, способной к рефлексии и обладающей свойствами «квазиличности». Приводится типология глобальных трендов, выделенных российским социологом О. Якимовой и связанных с ускоренным развитием искусственного интеллекта, расширением возможностей его применения в различных общественных сферах, профессиональной деятельности и повседневной жизни. Делается вывод о том, что указанные тренды основаны на взаимосвязи ключевых процессов развития искусственного интеллекта за счет технологических инноваций, стимулирующих появление новых профессиональных компетенций, что приводит к росту социальных изменений и влечет за собой гуманитарные последствия в современном обществе.

Ключевые слова: искусственный интеллект; нейронные сети; квазиличность.

Современный технократический бум, связанный с интенсивным развитием искусственного интеллекта (прежде всего – генеративных нейросетей вроде ChatGPT), несомненно, порождает вопросы о дальнейших перспективах и направлениях развития информационных технологий и их постоянно возрастающей роли в жизни современного общества. Искусственный интеллект (далее – ИИ) сегодня выступает как многозначное понятие, объединяющее в себе комплекс проблем информатики, кибернетики, математического моделирования и программирования, связанных с имитацией человеческого мышления и поведения (от автоматов до роботов), автономии самоорганизующихся информационных систем, которые не требуют специального контроля и вмешательства человека, а также переноса индивидуальных личностных качеств в виртуальную среду для создания «цифрового клона» конкретного человека. Разработки в сфере ИИ начали осуществляться с 1960-х гг., взяв за основу образцы автоматических устройств, способных осуществлять массовый сбор данных, их обработку и анализ, репрезентацию различных фрагментов реальности в виде детализированных и все более сложных моделей, применение автоматических систем принятия решений и проработки условий их реализации.

Возможность обучения считается одной из главных преимуществ ИИ перед традиционными алгоритмами, поскольку он не программируется в привычном смысле этого слова, но способен развиваться самостоятельно, без участия со стороны программиста или регулятора систем. При этом сущность обучения ИИ заключена в развитии способности к выявлению, распознаванию и постепенному воспроизведению образов. Это позволяет более или менее детально выявить, спроектировать и воспроизвести общие тенденции и структуры в больших массивах данных, служащих рабочим материалом для разработки модели процессов или событий, а выводимые из этой модели следствия составляют основу для прогнозов с вероятностными оценками. При наличии достаточного количества вычислительных ресурсов, детализированных алгоритмов по извлечению характерных признаков и доступа к достаточно большому количеству образцов система ИИ способна смоделировать даже особенности индивидуальной манеры и творческих особенностей какой-либо выдающейся личности в сфере искусства и создать новые произведения, например, в стиле Рембрандта или Баха. Технологическая основа для развития и совершенствования интеллектуальных систем разработана уже в

достаточной степени, чтобы они могли из практически любого вида человеческих навыков извлечь их неявные правила и переопределить в качестве упражнения по распознаванию и воспроизводству образов – даже таких, которые, казалось бы, в наибольшей степени зависят от «духовности» или интуиции.

Современное состояние интеллектуальных информационных систем и их объединение в едином концептуальном поле по разработке ИИ позволило бы выйти на качественно новый уровень, в бесконечный процесс совершенствования не с целью имитации возможностей человека, но создания чего-то более оригинального в логике постгуманизма. Однако, если автономные системы ИИ разовьют свои собственные критерии для оценок и логики обоснования решений, подключаться к ним и работать с ними не только отдельным людям, но всем социальным институтам и инфраструктурам станет достаточно затруднительно. Как считает футуролог Г. Леонгард, мы уже достигли точки, когда ориентированные на человека политика и стандарты, цифровая этика, социальные контракты и глобальные соглашения о гуманизации этих экспоненциальных технологий столь же важны, как и договоры о нераспространении ядерного оружия. Во многих отношениях это уже актуально, поскольку алгоритмические системы, которые способны непрерывно оптимизироваться в рамках предоставленных им наборов начальных параметров, создают некоторые разработки (например, инженерно-конструкторские и дизайнерские решения) совершенно нетривиальным образом [2].

Так возникает специализация и выделение современных направлений развития технологий ИИ в таких областях, как управление знаниями и «онтологический инжиниринг», Semantic Web, различные логики (четкие, нечеткие, дедуктивные, индуктивные, абдуктивные, дескриптивные, темпоральные, пространственные и т.д.), искусственные нейросети, байесовские сети и генетические алгоритмы (Machine learning в узком смысле), компьютерная лингвистика, семантический анализ текстов естественного языка, image processing как «техническое зрение» (семантический анализ изображений), speech processing – семантический анализ речевых сообщений, многоагентные системы, гибридные интеллектуальные системы и пр.

Особая роль разработчиками ИИ отводится нейронным сетям – автономной интеллектуальной системе, имитирующей работу нервных клеток живого организма на основе модели из соединенных и взаимодействующих между собой искусственных нейронов. Каждый из них обладает набором входов, передающих сигналы разной интенсивности, «телом», состоящим из сумматора и функции активации, и единственным выходом, интенсивность сигнала которого составляет результат обработки входных сигналов. Как правило, искусственные нейросети моделируются из нескольких слоев нейронов, на первый из которых передаются входные данные, затем используется один или несколько скрытых слоев и, наконец, выходной слой показывает результат. Например, для распознавания объекта могут выделяться 100 его отличительных признаков, которым будут соответствовать 100 нейронов входного слоя, затем последуют несколько скрытых слоев нейросети, а выходной слой может содержать лишь 10 нейронов.

Такая схема компоновки упрощенных вычислительных блоков открывает большие возможности для обработки данных, особенно при использовании специальных методов обучения – с «учителем» (образцом для обучения), без него и с подкреплением. В первом случае в сеть на вход подаются данные из некоторого обучающего набора (признаки объекта, которые преобразуются в интенсивность сигналов входных нейронов). Затем сигналы суммируются и активируют скрытые нейроны следующих слоев до тех пор, пока не будет достигнут выходной слой нейросети. Сигналы нейронов можно рассматривать при этом как ответ на определенный вопрос – например, о принадлежности изображения какому-то классу.

Если параметры нейросети подобраны верно, то после обработки достаточного количества обучающих образцов она становится способна классифицировать незнакомые объекты. Обучение без «учителя» строится таким образом, что при поступлении нового набора данных сеть пытается найти наиболее соответствующий ему из ранее обработанных и объединяет свое

представление о целой группе образцов в виде определенного кластера. Идентификация уникального образца потребует создания нового кластера, в котором будут накапливаться данные для последующей обработки, сравнения, выделения уникальных признаков и дальнейшей оценки структуры данных. Подобный способ обучения используется для поиска зависимостей в больших объемах необработанных и несистематизированных данных, как в медицине, где нужно учитывать индивидуальные параметры каждого пациента для постановки диагноза и отслеживания его состояния в связи с приемом определенных лекарств.

Обучение с подкреплением на основе алгоритмов «глубокого обучения» (deep learning) применяется для создания программ «интеллектуальных агентов», способных самостоятельно принимать решения в некотором замкнутом окружении – таких, как автопилоты, системы управления дронами и манипуляторами промышленных роботов, разнообразные боты. Такая сеть учится использовать некоторую функцию «вознаграждения» (quality function), которая отображает параметры среды и возможные действия на некоторую величину «награды». Особенно хорошие результаты агенты показывают в компьютерных играх, где в условиях полностью виртуальной среды они быстро и экономично обучаются, играя сами с собой. В физическом мире выстроить параметры взаимодействия со средой и смоделировать развитие ситуации гораздо сложнее, но прогресс и в этой области уже настолько очевиден, что вызывает беспокойство у многих людей. Например, С. Хокинг считал, что нейросети могут в ближайшем будущем заменить людей во многих сферах деятельности, не связанных с творческим и интеллектуальным трудом, что потребует пересмотра экономического уклада современного общества [3].

Между тем, уже существуют рекуррентные нейросети – особый класс нейросетей, в котором предыдущие состояния системы используются для вычисления следующего за счет знания не только правил преобразования входного сигнала в исходящий, но и эволюции состояний. Использование при этом «краткосрочной» и «долговременной» памяти позволяет в соответствии с состоянием сети применять части «воспоминаний» при последовательной обработке информации. Этот вид сетей может опираться на временную последовательность входных сигналов, находить закономерности на основе полученного опыта, а при возбуждении входного слоя выдавать уже собственную последовательность данных. Так возникает возможность моделирования творческих способностей, аналогичных человеческим – например, в сочинении музыки, написании романов или живописи, причем машине удастся успешно имитировать индивидуальный стиль того или иного художника, писателя или композитора.

Но можно ли считать такой «продвинутый» ИИ аналогичным человеческому и в чем его базовое отличие от «естественного»? Для ответа на этот вопрос в качестве отправной точки можно взять известный тест А. Тьюринга, в котором человеку предлагается определить, взаимодействует ли он с искусственным интеллектом или с другим человеком. В случае успешного прохождения этого теста «умным» механизмом или техническим устройством, способным имитировать свойства человеческого интеллекта, можно считать, что найден важный критерий для оценки работы такого устройства. Однако даже в этом случае вывод, что создана некая «искусственная личность», во многом аналогичная человеческой, будет явно поспешным [5].

В этой связи британским философом и логиком Дж. Сёрлем был введен термин «сильный искусственный интеллект», который обозначает систему, не просто способную пройти тест А. Тьюринга, но и «осознать» используемые ею алгоритмы. В таком случае возникает эффект рефлексии и может сформироваться некая «квазиличность», которая не просто пользуется словами по определенным правилам, реагируя на вопросы, но понимает слова и то, как именно ими пользоваться в соответствии с правилами литературного языка. Однако даже если в тесте А. Тьюринга уже невозможно отличить, кто с нами общается (человек или машина), из этого вовсе не следует, что мы имеем дело с полноценной личностью или носителем разума, аналогичного человеческому.

В итоге российский исследователь И. Ларионов выделяет два критерия, которые могут служить ориентирами как для технических специалистов, так и для гуманитариев при обращении к теме ИИ. В качестве таких критериев выступают: 1) «полноценная автономия (удаленность от разработчика/оператора или эффективное функционирование без постоянного их вмешательства не могут быть критериями автономности); 2) интенциональность (не как психологический феномен воли и/или намерения, но как процесс осознания и оценки ситуации с применением определенной ценностной парадигмы)». Без выполнения данных критериев, по мнению Ларионова, ИИ не может считаться «самостоятельным искусственным агентом», т.е. автономным социальным актором, действующим на основе самостоятельно принятых решений [1].

Однако ускорение процессов развития ИИ сегодня таково, что уже в ближайшем будущем мы можем ожидать прорыва в этой области, позволяющего говорить о появлении нового типа личности как продукта и эффекта социальных и технических инноваций. В связи с этим российский социолог О. Якимова, редактор спецвыпуска издания «Социодиггер», посвященного ИИ, на основе материалов данного выпуска выделила семь развивающихся трендов, связанных с реакцией общества на внедрение эффективных инструментов ИИ в повседневную жизнь и профессиональную деятельность людей [4].

В качестве первого тренда выделено развитие возможностей генеративных нейросетей в области создания осмысленных текстов большого объема и поддержки диалога с человеком на естественном языке благодаря прогрессу функционального ИИ, решающего узкоспециализированные задачи и успешному применению методов обработки сложноструктурированных данных (изображений, видео, текстов).

Вторым трендом названо создание «искусственной социальности» на основе трансформации человеческой деятельности, социальных взаимодействий и повседневности в целом, когда ИИ выступает в качестве субъекта коммуникации и требует от нас соответствующей подготовки – повышения цифровой грамотности, понимания ограничений в использовании ИИ, а также его технических характеристик с целью смягчения последствий неприемлемого применения и злоупотреблений.

Третьим трендом считается смена парадигмы образования, особенно в сфере социально-гуманитарных дисциплин – для оценки получаемой информации (знаний) в пользу обучения и проверки умения мыслить, что требует развития навыков анализа найденных самостоятельно или с помощью инструментов ИИ данных, обучения процессу поиска информации и его уточнению; понимания адекватности получаемых в результате запроса в интернете данных и пр.

Четвертый тренд связан с трансформацией речевых практик и традиций в науке и журналистике, поскольку повсеместное распространение GPT-моделей может привести к подстройке языковых процессов под логику нейросетей, релятивизации научных и публицистических текстов, уменьшению в них числа сильных утверждений, снижению качества логической аргументации и т.п.

Пятый тренд проявляется на рынке труда в связи с изменениями в профессиональной деятельности многих специалистов, для которых важными будут навыки работы в тандеме с ИИ или даже в качестве ассистента для него. Отдать ключевые когнитивные задачи на аутсорсинг ИИ станет возможно в сферах образования, медицины, политики, медиа и индустрии развлечений благодаря прогрессу технологий ИИ, меняющих условия труда, обеспечивающих его цифровизацию и компьютеризацию, способствующих его «программированию» и «модернизации».

Как следствие, проявится шестой тренд в связи с влиянием ИИ на социальное неравенство при увеличении цифрового разрыва (digital divide) в навыках между теми, кто применяет и не применяет ChatGPT при выполнении профессиональных задач, поскольку использование

ИИ серьезно уменьшает время их решения, а также повышает качество работы и удовлетворенность от нее. В итоге научные коллективы, учебные организации или даже целые страны, которые откажутся от использования генеративного ИИ, существенно отстанут от тех, кто будет активно ее использовать.

Наконец, седьмой тренд связан с распространением массовых негативных представлений и стереотипов за счет «порочного круга» обучения нейросети на «массовых представлениях» и генерировании ею еще большего количества представлений по аналогии с массовыми, которые выступают в качестве обучающей основы для нее же на новой итерации. Вследствие этого генеративные нейросети в своих выдачах поддерживают стереотипные образы – в частности, в сфере восприятия некоторых профессий.

Таким образом, указанные тренды основаны на взаимосвязи ключевых процессов развития ИИ за счет, во-первых, технологических инноваций, стимулирующих, во-вторых, появление новых профессиональных компетенций, что приводит к росту, в-третьих, социальных изменений и далее, в-четвертых – гуманитарных последствий в современном обществе и культуре. И это можно рассматривать как новый вызов, с которым столкнулось человечество на очередном витке научно-технического прогресса.

Библиографические ссылки

1. Ларионов И. Нейросети в 2023 году рождение искусственной личности? [Электронный ресурс] // Социодиггер. Т. 4. Вып. 7-8 (27), 2023. URL: <https://sociodigger.ru/articles/articles-page/neiroseti-v-2023-g-rozhdenie-iskusstvennoi-lichnosti> (дата обращения: 27.03.2024).
2. Леонгард Г. Технологии против человека. М.: Издательство АСТ, 2018. 320 с.
3. О'Коннелл М. Искусственный интеллект и будущее человечества. М.: ЭКСМО, 2019. 272 с.
4. Якимова О. Жизнь с ИИ: от взаимодействия к взаимозависимости [Электронный ресурс] // Социодиггер. Т. 4. Вып. 5-6 (26), 2023. URL: https://sociodigger.ru/articles/articles-page/zhizn-s-ii-ot-vzaimodeistvija-k-vzaimozavisimosti?utm_medium=email&utm_source=UniSender&utm_campaign=302232594 (дата обращения: 27.03.2024).
5. Turing A. Computing Machinery and Intelligence // Mind, 1950. Vol. 59. No. 236. P. 433-460.