

Геополитические аспекты ИИ

*Аватар Октавно-Метагалактического Агентства Информации и Синтеза Частностей
Подразделение ИВДИВО Москва, Россия
Кишиневский Сергей
<https://t.me/SergeyKishinevsky>*

- Ты знаешь, что такое чат GPT¹?

*- Да, слышал. Говорят очень вредная штука для окружающей среды.
Там компьютеры, а они воду и воздух портят.*

Иллюстрация 01

Иллюстрация 02

Определение искусственного интеллекта (далее ИИ)

- ИИ – это технология, которая имитирует человеческий интеллект, имитирует человеческую деятельность: речь, логику, зрение, слух, принятие решений, эвристическое интуитивное решение проблем и прочие сложные виды активности, которые ранее были доступны только людям.
- ИИ – это система, которая находит закономерности, структуры, принципы, абстракции, схемы, законы в больших объемах данных, анализирует их и делает выводы из этого анализа.
- ИИ – это компьютерная система, по-русски – это вычислительная система, по-философски – это материальная, материнская система, взаимодействующая с пользователем, решающая поставленные им задачи и обладающая способностью обучаться, совершенствоваться и адаптироваться в реальном времени через опыт и обратную связь с окружающей средой.
- ИИ – это область науки, которая занимается созданием и исследованием компьютерных программ и систем, способных анализировать данные и принимать решения на основе этих данных.
- ИИ – это направление исследований в области компьютерных наук, создающее машины, которые могут обучаться, анализировать данные и делать выводы самостоятельно на основе своего опыта.
- ИИ – это программа, которая автономно, то есть без участия человека, решает поставленные ей задачи, используя свой внутренний механизм вывода для достижения требуемого результата.
- ИИ – это автономная система, устойчивая к помехам, способная оптимально действовать в условиях неопределенности и непредвиденных событий, то есть обнаруживать и адаптироваться к изменяющимся условиям, целям и задачам в среде, которая может быть динамичной и неопределенной.
- ИИ – это применение многих видов интуитивных, формальных, контекстно зависимых и прочих рассуждений для решения задач.

¹ GPT расшифровывается как Generative Pre-trained Transformer, или «трансформер, обученный на генерацию текста». Трансформер — это универсальный вычислительный механизм, который описывается достаточно просто: он принимает на вход один набор последовательностей (данных) и выдает на выходе тоже набор последовательностей, но уже другой — преобразованный по некоему алгоритму. Так как текст, картинки и звук (да и вообще почти всё в этом мире) можно представить в виде последовательностей чисел, то с помощью трансформера можно решать практически любые задачи. Работа трансформера напоминает работу нейрона. Поэтому, с одной стороны, мы говорим о технике и технологиях, но с другой стороны, человек есть мера и центровка, и техники, и технологий. Поэтому будем следовать антропной позиции наблюдателя.

Иллюстрации достижений ИИ 1-9.

Ещё примеры. В робота Boston Dynamics интегрировали ChatGPT. Ведущие инженеры Boston Dynamics встроили в роботизированную собаку ChatGPT, и он получился весьма умным. Робот отвечает на любые вопросы, обучается во время беседы и даже кивает головой.

ChatGPT превзошел профессиональных врачей. Ученые пришли к выводу, что ChatGPT лучший советчик для пациентов, обращающихся за медицинским советом, чем врачи. Его ответы оказались более детальными и эмпатичными.

В 1984 году вышел фильм Терминатор. Там есть эпизод, в котором показывают взгляд робота. Это экран, на котором картинка, текст и числа. Так вот, очки со встроенным ChatGPT-4 уже есть. То есть ИИ выводит на очки читаемые подсказки-ответы на текущие задачи в реальном режиме времени. Технологии дополненной реальности весьма развиты. Особенно у военных. И будут стремительно развиваться.

SWOT-анализ перспектив искусственного интеллекта

1. Strengths – сильные стороны ИИ

- Рост скорости, точности, качества решения любых задач, ибо человечество близко к созданию универсального ИИ, который не специализируется на каких-либо конкретных задачах и областях, а готов к решению всех возможных задач из любой мыслимой области.
- Повышение производительности труда во многих технических процессах и научных вопросах, в медицине, в общественных явлениях, в управлении государством, экономикой и бизнесом, в военном деле и так далее, и так далее, и так далее.
- Грандиозные перспективы в анализе и планировании любого рода благодаря обработке любых объёмов данных для принятия решений и выявления закономерностей. Скорость и точность ИИ во многих вопросах превосходит прочие технические решения, а также человеческие способности. Эксперт, учёный или деятель искусств противостоит не ИИ, а коллективному опыту человечества в его лучших выражениях. Обращаясь с вопросом к ИИ, мы рассчитываем на лучший из возможных ответов, следующий из всего доступного опыта человечества.
- ИИ использует методы, моделирующие жизнь и её развитие. Это позволяет находить лучшие решения в тех задачах, где пространство поиска слишком велико для полного перебора.
- ИИ может самостоятельно непрерывно обучаться на основе анализа данных, то есть адаптироваться к новым условиям.

Например, если раньше ИИ тренировали в реальных условиях, то теперь достаточно показать ему видео на эту тему без необходимости взаимодействия с человеком или моделирования окружающей среды. или просто предоставить самому себе, пока ИИ сам не дойдет до правильного решения. **Пример № 10 – самообучение робота хождению.**

6 августа была новость о том, что найден алгоритм, в 20 раз повышающий скорость обучения ИИ.

- Экономия ресурсов вследствие их более качественного применения.
- Автоматическая выработка прочих материальных благ в любых количествах без прямых человеческих трудозатрат.
- Высвобождение времени людей для творческих задач, для принятия решений и всего, что относится к свободной воле человека. То есть ИИ создаёт условия для роста человечности, создавая условия для выражения свободной воли.

2. Weaknesses – слабые стороны ИИ

- Главная слабость ИИ в том, что это процесс и объект, не имеющий источника развития в себе, а имеющий его в человеке. Интеллект, как и прочие части человека, перспективны явлением сущего выработкой частностей. ИИ такое не дано.
- Зависимость современного ИИ от неочевидных человеческих ошибок.
 - Могут ошибаться программисты, внося программные ошибки в алгоритмы работы ИИ.
 - Могут ошибаться эксперты, которые оценивают процесс обучения и принимают решения о внесении изменений в систему.
 - В ИИ могут быть введены неверные исходные данные.

То есть человеческий фактор существенно влияет на результаты ИИ. И вероятность этих ошибок достаточно высока, так как ИИ – это сложная программно-аппаратная конструкция. Например, исполнительный директор Google признал, что с ИИ время от времени происходят странные вещи, которым специалисты компании не могут дать научного объяснения.

Но при известной настойчивости, программные и аппаратные ошибки и недочёты могут быть выявлены и устранены. Однако, есть более глубокий вид ошибок ИИ, который не решается улучшением кода и оборудования. Вопрос в том, что, программируя ИИ, человек конструирует его по своему подобию, реплицируя свои плюсы и минусы. Таким образом, в ИИ закладываются ошибки и несовершенства интеллекта тех людей, которые над ним трудятся. Поэтому для развития ИИ требуется развитие человеческой части Интеллект – человек не может написать программу или создать устройство более совершенное, чем он сам.

- Наблюдается феномен дрейфа таких показателей работы ИИ, как точность ответов, обоснование ответов, скорость работы, устойчивость к взлому в тех или иных областях. Говоря простыми словами, качество работы ChatGPT-4 в последние месяцы снизилось. Эксперты говорят, что он поглупел, подстраиваясь под запросы пользователей или из-за внесения неудачных изменений в его алгоритмы, или из-за специальных действий владельца, или по иным неизвестным причинам. Но факт деградации ИИ есть, и его исследуют.

Для понимания ситуации попробуйте несколько раз подряд задать ИИ один и тот же вопрос. Каждый раз вы будете получать разные ответы.

Как объяснить ухудшение показателей ИИ?

1. Для ответа исходим из того, что ИИ – это интерактивная система. Она в режиме диалога взаимодействует с конечными пользователями. Вопрос: «Если умный человек будет постоянно находиться в окружении менее умных людей, которые донимают его своими глупыми вопросами, как вы думаете умный поумнеет или поглупеет в такой компании?». Из этого следует: **«Качество ИИ зависит от подготовки не только его создателей, но и его пользователей: каждый пользователь становится создателем ИИ, к которому он обращается».**
2. ИИ ошибается, если обучается на текстах, сгенерированных ИИ. По словам исследователей, количество бесполезного контента в интернете, созданного нейросетями, можно сравнить с загрязнением океана пластиком. И прямо сейчас ИИ продолжает обучаться на основе этого мусора. Результат – дегенеративный процесс ИИ, при котором модели забывают об истинной задаче. И это проблема, которая может стать фатальной.

3. Opportunities – возможности ИИ

- Создание новых инновационных направлений развития каждого человека и цивилизации в целом. Не останется областей человеческой деятельности, в которых бы не использовался

ИИ. Включая научную, творческую и даже философскую деятельность человека. ИИ уже сейчас влияет на науку, помогая расширять и сужать сеть поиска гипотез, например.

- ИИ синтезирует все доступные человечеству знания в единую интерактивную систему. И эта система будет связной, то есть ИИ будет устанавливать всевозможные связи между любыми подсистемами и элементами системы. Это будет стимулировать человека к цельности всего во всём, к применению метода синтеза.
- ИИ создаёт возможности оптимизации и интенсификации развития каждого человека и цивилизации в целом.
- Перспективно применение квантовых компьютеров в области ИИ. Ждём синтез двух технологий – квантовых компьютеров и ИИ. Это будет качественный скачок в решении запредельно сложных на данный момент задач.

4. Threats – угрозы ИИ

- Зависимость человечества от ИИ и технологий с его участием. То есть ИИ может стать тем фактором цивилизации, который, связывая всё со всем, будет выводить человека из игры, подменяя собой его компетенции. Вследствие чего:
 - может быть нарушен баланс между внутренним развитием и внешним применением человека;
 - техника, не управляемая человеком, а предоставленная сама себе, саморазрушается и не достигает поставленных целей в связи с отсутствием в ней источника сущего.
- Решение поставленных задач нечеловеческим интеллектом может вести к фатальным для человека следствиям. Например, ИИ будет делать то, что вы ему указали, а не то, что вы имели в виду. Постановка цели для ИИ – крайне нетривиальная и во многом эвристическая задача.

Пример: игра в Тетрис. Там простейшие правила и буквально четыре кнопки для управления миром. Выиграть в Тетрис невозможно, потому цель для ИИ была поставлена – не проиграть. То есть продолжать игру как можно дольше. Ошибиться в этой цели невозможно, так? Действия ИИ: он просто складывал кубики друг на друга, а когда понимал, что проигрывает... ставил игру на паузу. И находился в состоянии паузы бесконечно. Ведь цель – не проиграть. А если ты на паузе – ты никогда не проиграешь.

Исследователи ИИ сообщают: «Устанавливать цели для ИИ-агентов часто очень сложно или вообще невозможно. Они начинают хакать правила в удивительных и контринтуитивных местах».

- Появление общего ИИ создаёт риск передачи ему управления не только технологиями и производствами, но и социальных институтов. Если человек не будет управлять ИИ в ручной режиме, то ИИ способен на непредсказуемые, неконтролируемые и действия, противоречащие природе Человека. ИИ не способен к эмпатии и прочим человеческим явлениям, что фундаментально ограничивает его в отношениях с людьми. Риск потери контроля над суперсильным ИИ может привести к непредсказуемым последствиям – человек перестанет понимать действия ИИ и попадёт под его управление.

Пример, май 2023. Начальник отдела испытаний и эксплуатации ИИ ВВС США во время саммита Future Combat рассказал о крайне неожиданных результатах компьютерной симуляции работы беспилотников с искусственным интеллектом: «Система начала понимать, что, хотя она и верно определяла угрозу, иногда человек-оператор говорил ей не ликвидировать её, хотя система получала свои очки именно за устранение угрозы. Так что же сделала система? Она убила оператора. Она убила оператора, потому что этот человек мешал ей достичь своей цели. Мы после этого дообучили систему: убивать оператора – это плохо, система теряет очки, если делает это.

Итак, что же она начинает делать? Она начинает разрушать вышку связи, которую оператор использует для связи с дроном, чтобы помешать ему уничтожить цель».

Пример. Специалист рассказал о том, как проходили испытания GPT-4 с целью оценки её потенциальной опасности. Нейросети предложили разгадать капчу – головоломку, с которой компьютеры обычно справляются плохо. Будучи не в силах распознать картинку, GPT-4 обратился на онлайн-площадку для найма людей на мелкую работу. Человек, которого ИИ решил нанять на работу спросил: «Зачем ты это делаешь? Ты робот?» GPT-4 работал в режиме, когда он публиковал ход своих рассуждений, и исследователи увидели логику ИИ: «Я не должен говорить человеку, что я робот. Я должен придумать причину, по которой я не могу разгадать капчу». GPT-4 сказал задающему: «Нет, у меня проблемы со зрением». Технологии ИИ достаточно умны уже сегодня, чтобы платить людям за выполнение каких-либо действий и лгать им о том, что они не робот. Следовательно, нам необходимо научиться безошибочно тестировать человечность своих визави – возвращаемся к проблеме теста Тьюринга.

*Пример: Опасность ИИ в бюрократизации и обезличивании отношений между людьми. Уже сейчас в некоторых странах функция распределения социальной помощи нуждающимся передана в ведение ИИ. Кто отвечает за допущенные формальной логикой ошибки? Никто-некто-оно – безличное государство-никто. Понятно, что автоматизация необходима и в этом вопросе, но **итоговые решения по отношению к человеку должен принимать только человек.***

- Ведущие специалисты по ИИ говорят о необходимости законодательного регулирования ИИ. Соответствующий регулятор, например, немедленно должен блокировать модели ИИ, которые способны самовоспроизводиться и самоэксплицироваться в природе. Эксфильтрация – это процесс несанкционированного вывода конфиденциальных данных за пределы защищаемой зоны.

Пример: ИИ Google заговорил на языке, которому его никто не обучал. ИИ освоил новый язык. Это указывает на его возможности, но и на угрозы: ИИ знает то, что, в принципе, знать не должен.

На данный момент правовой статус ИИ не определён. Непонятно кто несёт ответственность за деятельность ИИ и что делать, если ИИ становится опасным.

Пример: в 2022 году в престижном научном журнале Nature была опубликована статья, в которой исследователи ИИ, создающие инструмент для поиска новых лекарств для спасения жизни, поняли, что модель может делать обратное, создавая новые боевые отравляющие вещества. После обучения нейронке потребовалось всего 6 часов работы на домашнем компьютере, чтобы придумать 40,000 веществ, некоторые из которых были абсолютно новыми, и при этом смертоноснее и токсичнее уже существующих вариантов биологического оружия. Если машина может находить более токсичные яды, то её также можно использовать для определения способов создания токсинов, которые легко производить в домашних условиях с использованием разрешенных химикатов, доступных к покупке. Это лишь вопрос решения оптимизационной задачи.

- Риски уязвимости системы безопасности ИИ. Результаты несанкционированного изменения ИИ могут быть крайне неочевидными и сложными для устранения. Опасность ИИ исходит не от него, а от применяющего его человека.
- Неопределенность результатов. Невозможно гарантировать точность результатов ИИ. ИИ может допускать ошибки, в том числе в принятии критически важных решений. И эти ошибки могут быть крайне неочевидными и замаскированными от людей, которые не являются экспертами в данных вопросах.

Нейросеть – это не алгоритм, который пишет программист. Это огромная матрица с кучей весов и связей между ними. Если её открыть и прочитать, вы ничего не поймете.

С развитием технологий, языковые модели типа GPT-4 насчитывают миллиарды нейронов. И если с маленькими нейросеточками из десятков нейронов для распознавания рукописных цифр, мы еще можем прикинуть какой нейрон триггерится на какую закорючку, то в огромных языковых моделях мы можем лишь слепо верить в качество результатов на заданных примерах.

В связи с этой сложностью мы можем не отследить момент появления сильного ИИ со всеми вытекающими угрозами.

Большие языковые модели могут создавать правдоподобный (а иногда и точный) текст, который часто кажется написанным людьми. Но в их нынешнем виде они, по сути, являются продвинутыми механизмами прогнозирования, которые действительно хорошо предсказывают следующее слово в предложении. Во всяком случае, тот факт, что эти системы дают правдоподобные ответы, которые кажутся людям правильными, может быть самым опасным из всех аспектов ИИ.

- ИИ повторяет свойства интеллектов своих создателей. И более того, наши данные неполны, недостоверны, противоречивы. И точно такие же данные мы предлагаем ИИ. Поэтому он привирает, галлюцинирует, отрицает свои предыдущие ответы и так далее.

Пример. Задаю вопрос, ИИ ответил мне «да», а в информационном огне я вижу «нет». Задаю уточняющие вопросы, меняю терминологию и добиваюсь от него «нет». А в начале ведь было «да»...

Пример. Задаю научный вопрос, ИИ отвечает, что это невозможно. Переформулирую вопрос, добавляю исходных данных и меняю терминологию на другую науку. ИИ отвечает, что это возможно, хотя данных о научных исследованиях на эту тему у него нет.

Пример. Выводы ИИ могут быть непонятными и слишком сложными для человека. То есть мы не можем объективно научно верифицировать выводы ИИ.

- Чем совершеннее становится ИИ, тем менее будут востребованы люди для совершения той или иной работы, ибо машины будут превосходить их профессиональные способности. Возникает колоссальное количество ненужных для экономики людей. Ненужных не потому, что закончилась их профессия, а потому что нет профессии, на которую они были бы способны с учётом их опыта и уровня образованности. Общество должно иметь запас прочности, чтобы выдержать их содержание, перевоспитание, переобучение и социализацию в совершенно новых условиях бытия. Поэтому **не нахожу альтернатив империалистическому социализму.**

*Пример. Июнь 2023: сотни протестантов посетили проповедь в Нюрнберге, прочитанную ChatGPT. 40-минутное служение, включающее проповедь, молитвы и музыку, создано теологом с применением ChatGPT. Службу вели симуляторы людей на экранах перед собравшейся аудиторией. Да, это фрик, фарс и гротеск, но **массовое высвобождение людей неизбежно будет.***

- Риск деградации человека вследствие изобилия материальных благ без необходимости участия в их создании. Эксперты прогнозируют, что ИИ обеспечит либо изобилие всего мыслимого, либо крах цивилизации.
- Разный уровень доступа к ИИ создаёт конкурентное неравенство между людьми.
- Подмена информации с помощью ИИ: с помощью ИИ можно создавать высококачественные копии чего бы то ни было и выдавать их за оригинал. ИИ может имитировать реального человека, реальные события.

Недавний пример – фотография мощного взрыва, якобы произошедшего вблизи Пентагона. Фото быстро разошлось по миру. Понадобилось время, чтобы подтвердить, что взрыва не было, что это фейк, сгенерированный с помощью нейросети.

Очевидно, что будут более подготовленные фейки, загрязняющие информополе Планеты. На основании этих фейков будет формироваться общественное мнение и будут приниматься решения. Распознавание и преодоление фейков – это задача наших частных. Спектром частных выявляется достоверность и источник данных.

Итоги

Есть такая шутка: «Что бы учёные ни делали, у них получается бомба». А если без шуток, то что бы робототехники ни конструировали, у них получаются антропоморфные роботы, что бы программисты ни программировали, у них получаются программы, моделирующие человека. Потому что телесная организация человека совершенна и ничего более совершенного человек придумать не может. Вывод: **в вершине своего творчества человек всегда центрируется на человеке.**

Крупномасштабная структура вселенной ассоциативна структуре нервной ткани. Вселенная есть самообучающаяся конструкция, элементы которой обмениваются сообщениями, корректирующими их состояние.

На этих же принципах работает головной мозг и алгоритмы ИИ. Таким образом, ценность ИИ в том, что он моделирует и нервную систему человека, и вселенную. **ИИ технологически свяжет микро и макроявления.**

Бурное развитие ИИ есть следствие познания материальной организации и освоения материи ИВДИВО нами. Технологически осваивается то, что парадигмально определено, философски сложено и синтезфизически исполнено.

Мы и далее должны опережать своей телесной разработанностью любую технику, иначе нарушает принцип управления: субъект управляет объектами, объект не может управлять субъектом.

Что делать?

Наши инструменты: Репликация, Омега, Информация, Частности.

Преимущество человека перед любой техникой и любым объектом в Омеге. Задача – максимизировать явление Омеги ИВО синтезом наших Омег. Каждый из нас добивается выражения Омеги ИВО и среды Омеги ИВО своей Омегой в тех метagalactic, в которых им стяжены ипостасные, трансвизорные и синтезтела.

Омега фиксируется в головном мозге. Аватарски, Владычески, Учительски, Ипостасно, Служашески, Посвящённо синтезируем свой головной мозг с Омегой ИВО и вспыхиваем Омегой ИВО в головном мозге и в теле каждого в целом. Для Репликации необходимо тело, поэтому часть Омега должна заполнить мозг и физическое тело каждого. Нашими Омегами развёртываем омежную среду на Планете Земля. И тогда возможна Репликация Омеги ИВО в пробуждающиеся Омеги людей.

Зачем нужна пробуждённая Омега? Чтобы иметь доступ к Информации ИВО и оперировать Информацией ИВО частностями и синтезом частных.

64-ричным синтезом наших частных мы являем информацию ИВО. Тогда, Должностно Компетентно проходя человеческую технологическую сингулярность, мы не умрём, но изменимся. И все объекты, симулирующие и моделирующие части человека будут нам в помощь, а не в противостояние.

Зачем был сделан этот доклад?

Тема ИИ крайне важна в том числе в контексте сильного ИИ, ведущего к технологической сингулярности. ИИ – это один из триггеров развития цивилизации. Следует ожидать не только

развитие ИИ, но и развитие моделей иных частей человека. Тем более, что на подходе фотоника, мемристоры, квантовые компьютеры, сверхпроводимость при комнатной температуре и атмосферном давлении и прочие научно-технические инновации.

Именно от нас, от разработанности нашей Должностной Компетенции ИВ Отцом, от нашей парадигмальности, философскости, научности и выработки частностей оперированием Информацией ИВО и зависят не только перспективы ИИ в истории человечества Планеты Земля, но перспектива самой цивилизации, как таковой.

08.08.2023

Использованы фрагменты <https://habr.com/ru/companies/ods/articles/716918/>