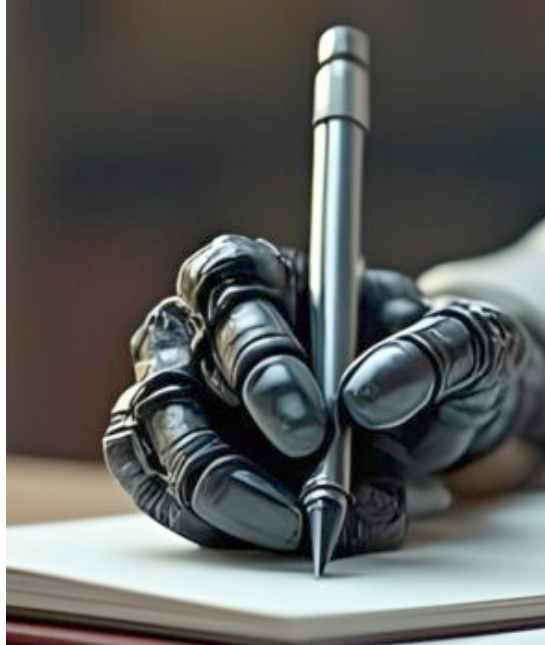




Эпоха разумных систем:

Сергей Марков о том, куда движется ИИ

Сегодня искусственный интеллект – зонтичный термин, охватывающий множество разрозненных технологий, разрабатываемых командами по всему миру для решения различных задач. На данный момент среди всего многообразия направлений выделяются 2 ключевых – генеративный и интерактивный (агентный) ИИ. И если первый сосредоточен на создании нового контента – будь то тексты, изображения или музыка, – используя мощные модели, обученные на обширных данных, то второй предназначен для принятия автономных решений, достижения сложных целей. При этом система агентного ИИ не просто генерирует контент, она разрабатывает стратегии, планирует и выполняет многошаговые действия с автономностью, выходящей за рамки простого распознавания образов, что делает ее незаменимой в условиях быстро меняющегося мира. Об основных характеристиках, примерах и областях применения генеративного и агентного ИИ, их взаимосвязи и потенциальном воздействии на будущее технологий и нашу жизнь рассказывает Сергей МАРКОВ – специалист в области искусственного интеллекта и машинного обучения, создатель шахматной программы SmarThink и основатель научно-популярного портала «XX2 ВЕК». В настоящее время он занимает пост директора по развитию технологий ИИ в «Сбербанке», где руководит исследованиями передовых нейросетевых моделей, таких как GigaChat, Kandinsky, SymFormer и др.



■ Генеративный искусственный интеллект – довольно хорошо знакомая большинству людей концепция. Он по-прежнему остается в тренде или это скорее вчерашний день?

– Его все еще нельзя назвать вчерашним днем: хотя первые генеративные модели действительно появились довольно давно, однако реальный прорыв в этой области начался только с появлением нейросетевой архитектуры под названием «трансформер» в 2017 г. Она была разработана исследователями Google и стала впоследствии основой большинства современных генеративных моделей, включая GPT, BERT, T5, Claude, Gemini и др. Так что генеративный ИИ на основе больших трансформерных моделей с нами всего лишь 8 лет. И, конечно же, еще есть множество задач, которые пока не решаются так, как бы мы того хотели. Но вслед за системами генеративного ИИ появляются интерактивные системы. Они, в свою очередь, базируются на всех достижениях генеративных моделей, но обладают гораздо большими возможностями. Принципиальная разница между ними заключается в том, что если система генеративного искусственного интеллекта, как правило, порождает результат «за один присест», к примеру генерирует изображение или видео по текстовому запросу, то интерактивная решает задачи на нескольких этапах или множеством шагов. Система не просто создает контент, а выполняет последовательности действий, направленные на достижение поставленных целей. В процессе решения задач она может взаимодействовать с другими системами, пользователями, использовать внешние инструменты. Основным строительным блоком современных агентных систем являются так называемые LLM-агенты – то есть агенты на базе больших языковых моделей (Large Language Models). Первые многоагентные системы, опять же, появились довольно давно – еще в 1970-е гг., поэтому в агентности как таковой нет особой новизны, но появление больших языковых моделей придало многоагентному направлению «второе дыхание».

Таким образом, большие языковые модели из фундаментальных (или базовых) моделей стали фундаментальными агентами. По сути, модели специальным образом обучают для того, чтобы они могли не только генерировать тексты, но также включать в них специальные управляющие команды, позволяющие агентам решать сложные, многоэтапные задачи.

■ Иными словами, они думают, размышляют, планируют?

– В определенном смысле. Одна из их особенностей – способность строить цепочки рассуждений. Но это не единственная отличительная черта LLM-агентов. Обычно речь идет о наличии механизмов долгосрочной памяти, но опять же с возможностью работать с цепочками действий, то есть генерировать какие-то управляющие команды, вызывать внешние инструменты, функции. Это можно сравнить с умножением пятизначных чисел, которое можно, конечно, попытаться выполнить в уме, но обычно мы решаем подобные задачи с помощью ручки и бумаги или калькулятора. По сути мы так или иначе используем какие-то внешние по отношению к нашему мозгу инструменты. Примерно этому сегодня обучаются и генеративные модели, что затем позволяет создавать из них агентные системы.

■ Считаете ли вы, что нынешний бум, подпитываемый большими языковыми моделями, принципиально отличается от предыдущих? Или мы снова рискуем столкнуться с «переохлаждением» темы из-за завышенных ожиданий?

– Я с некоторой долей иронии отношусь к популярной концепции «зим и весен» искусственного интеллекта. Очень часто те периоды, которые принято считать провальными, на самом деле характеризовались довольно высокой исследовательской активностью. Скажем, 1970-е гг. многими авторами считаются классической «зимой» ИИ. Но давайте разберемся, что было сделано в то время. А ведь именно тогда созданы первые сверточные нейронные сети, появление которых привело впоследствии к революции глубокого обучения; произошли довольно серьезные прорывы в решении задач распознавания образов, в частности в оптическом распознавании символов (Optical Character Recognition, OCR). Команда SRI (Стэнфордского исследовательского института) развивала успех, достигнутый в исследовательской программе MINOS, наследниками которой во многом являются современные системы распознавания текстов. За MINOS последовали другие успехи в обработке естественного языка – в SRI были получены переломные для своего времени результаты в распознавании речи, понимании диалогов и даже в области ИИ для управления роботами. Так что много нейросетевых исследований своими корнями уходит в то время, несмотря на то, что к этой области не проявлялось такого пристального внимания, как

в предыдущие годы. Правомерно ли считать это «зимой»? Отчасти лишь в том смысле, что общество обратило свой более пристальный взор на другие направления.

■ Но сегодня мы наблюдаем беспрецедентное внимание к искусственному интеллекту не только со стороны общества, но и ключевых институтов – финансовых, экономических, законодательных, а также колоссальные инвестиции в эту сферу. Не приведет ли этот ажиотаж, огромные финансовые вложения к новому витку застоя ИИ, когда неизбежно разочарование от неспособности технологии мгновенно оправдать все прогнозы? С другой стороны, существует реальная опасность того, что в погоне за быстрыми результатами и коммерческой выгодой человечество может упустить из виду вопросы безопасности и контроля над системами ИИ.

– Я не очень верю в застой – ведь сегодня системы ИИ уже стали частью повседневных сервисов и продуктов. Даже если не брать в расчет огромные массы людей, использующих на ежедневной основе сервисы, подобные ChatGPT, DeepSeek, GigaChat и т.д., системы ИИ являются незримой частью более чем привычных вещей. Любой поисковый запрос в Интернете, любая фотография, которую вы делаете при помощи современного смартфона, любой голосовой интерфейс – «под капотом» всего этого современные модели машинного обучения. Конечно, определенная переоценка и связанная в этом рыночная коррекция в той или иной области вполне возможна и даже, вероятно, неизбежна, но в глобальном смысле остановка прогресса в этой области вряд ли реальна.

Тема же рисков, связанных с развитием технологий, всеобъемлющая. Однако я с большой иронией отношусь к популярным страхам, связанным с искусственным интеллектом, сюжетам вроде тех, которые обыгрывались в фильме «Терминатор», или к сценарию так называемого «интеллектуального взрыва», когда самосовершенствование систем ИИ ведет к быстрому появлению сверхчеловека, неподконтрольного человеку. Есть вполне объективные причины считать, что такого рода прогнозы не сбудутся. Но это не значит, что к ним нужно относиться легкомысленно, пускать дело на самотек. Безусловно, есть объективные риски, исходящие от технологий ИИ, и они лежат совсем не в той области, которой общество обеспокоено. Если

посмотреть шире на внедрение различных алгоритмов в принятие решений, то можно заметить, что в значительной степени риски обусловлены технологическими и социальными причинами. Пример – система распознавания лиц, позволяющая решить множество проблем, связанных с безопасностью, созданием более удобных сервисов, те же биометрические системы, облегчающие использование сервисов и пр. Так, ее внедрение в аэропорту может сократить время прохождения контроля или ее применение обеспечит вход в квартиру без ключа. Однако в свое время американская исследовательница Джой Буоламвини из Медиалаборатории MIT провела любопытное исследование, посвященное системам распознавания лиц, чтобы оценить их точность при работе с лицами людей различного пола и с различным цветом кожи. Модели показали высокую точность в отношении светлолицых, но значительно хуже для темнокожих; кроме того, выяснилось, что лица женщин распознаются хуже, чем лица мужчин. История вполне объяснима: дело в дисбалансе примеров в обучающих выборках, которые использовались при создании систем. Но если представить себе, что подобная система распознавания лиц внедрена в реальные сервисы, то к чему это может привести? К систематическому неравенству в доступе к продуктам и услугам. А ведь речь может идти, скажем, о совершении платежа, регистрации в аэропорту или о каком-нибудь кредитомате. Получается, что большая социальная группа людей будет систематически сталкиваться с ограничениями, тратить больше времени и усилий для доступа к тем или иным услугам. Вот реальная история, которая началась в 1990-е гг. Сотрудники британской почтовой службы стали получать уведомления о недостачах в их почтовых отделениях и были вынуждены покрывать их из своих доходов. И это продолжалось много лет. Лишь совсем недавно работникам удалось в суде отстоять свою правоту и доказать, что причина была в дефекте информационной системы. Работники почты много лет добивались ее аудита, но компания всячески открещивалась от этого. Если оценить эту ситуацию с позиции конкретных служащих, то для них все было очень плохо – некоторые лишились работы, другие попали в тюрьму и даже есть подтвержденный случай самоубийства. Вот еще один вполне вероятный сценарий – отказ банка в кредитовании из-за банальной ошибки при вводе данных о клиенте, например даты рождения. По факту, как можно этому противостоять? Система защищена коммерческой тайной, доступа

к ней нет, как и возможности проверить правильность ее функционирования. Более того, пока не существует цифровых адвокатов, организаций или платформ, помогающих защищать права человека в ситуациях ошибок в применении алгоритмов. Если рассмотреть этот эпизод с точки зрения степени его влияния на чью-то жизнь, то такое решение вполне сопоставимо с вердиктом суда. Однако в суде, по крайней мере, есть возможность открытого доступа к собранным доказательствам, получения квалифицированной юридической помощи, состязательного процесса и пр. Простой пример: алгоритмы, управляющие транспортом, относятся к средствам и источникам повышенной опасности, поэтому предусмотрена серьезная процедура сертификации, подходы, минимизирующие риски, и они значительно выше, чем для традиционного транспортного средства. Видимо, стоит думать и о том, чтобы всякий раз при использовании ИИ были определены критерии, исключающие саму возможность появления любых ошибок, прозрачность алгоритмов и внедрение проверки данных. Такие меры помогут уменьшить количество неточностей и недоразумений. Также вполне возможно создание системы обжалования решений, осуществляемых на основе алгоритмов.

■ **Насколько оправдана и эффективна идея создания единого, универсального закона для ИИ, учитывая, что это, по сути, – инструмент, сравнимый с «умножением матриц», применяемый в областях с совершенно разной спецификой и рисками – медицине, финансах, транспорте? Не разумнее ли пойти по пути отраслевого регулирования, когда правила прописываются под конкретные случаи применения, а не абстрактную технологию? И вообще, не преждевременна ли сама попытка законодателя регулировать искусственный интеллект на нынешней стадии его развития, особенно в отдельно взятой стране? Речь идет, в частности, о разработке модельного закона по искусственному интеллекту как концепции, направленной на создание правовых норм и регулирования ИИ в России и Беларуси. Не рискуем ли мы, установив жесткие рамки, задушить инновации, отстать от глобальной технологической гонки?**

– Должен заметить, что в мире практикуется четыре подхода в этом направлении. Первый – американский, где не существует единого закона, регу-

лирующего работу такого рода систем. С моей точки зрения, его появление было бы немного странным. Давайте для наглядности заменим термин «искусственный интеллект» на «умножение матриц», что, конечно, будет не слишком точным, но аналогию, как мне кажется, позволяет почувствовать. Послушайте, как это звучит нелепо: «Закон для регулирования умножения матриц». Технологии ИИ достаточно универсальны и многообразны, они могут применяться в очень разных областях, каждая из которых обладает своей спецификой. Поэтому в США регуляторы пошли по пути создания рекомендаций и стандартов при разработке различных методов и технологий в конкретных областях применения. Кроме того, для отдельных сфер существуют отдельные законодательные акты на уровне штатов. Иными словами, нет единого закона.

Другой подход – ЕС, где действует так называемый AI Act, представляющий собой комплексный нормативный акт, направленный на регулирование искусственного интеллекта. Это гигантский документ, состоящий из нескольких сотен страниц, влияние которого на разработчиков эксперты оценивают как катастрофическое. Потому что если те, кто, например, создает большие языковые модели, будут исполнять все требования закона, то такие модели просто никогда не появятся. Так что сегодня для разработчиков больших языковых моделей, таких, как, например, Mistral, в Европе делаются специальные исключения. На мой взгляд, данный закон скорее используется как инструмент реализации принципа: «своим все, а остальным закон», чтобы выдвигать более жесткие требования к зарубежным компаниям, приходящим на европейский рынок. При этом, несмотря на всю многолетнюю бюрократическую волокиту, закон получился «сырым» и вызывает массу нареканий. Такое непродуманное закручивание гаек в условиях мировой гонки за развитием и обладанием технологиями в области искусственного интеллекта – «выстрел себе в ногу».

Есть и третий вариант, который условно можно назвать китайским. Он основан на принципах централизованного управления, государственного контроля и строгого соблюдения национальных стандартов безопасности и этики, при этом регуляторика представлена несколькими правовыми актами, такими, например, как «Меры по управлению службами генеративного искусственного интеллекта» (2023) или «Положение об администрировании информационных генеративных интернет-сервисов на базе глубокого обучения в Интернете» (2022).

Отметим, что здесь нет попытки создать единый закон для любых способов применения ИИ-технологий. Если внимательно изучить вышеупомянутые документы, то можно заметить, что в них очень много написано о том, к чему надо стремиться, создавая такие системы, какие существуют приоритеты, каким принципам необходимо следовать и т.д. При этом практически начисто отсутствуют какие-либо санкции для разработчиков, кроме возможной блокировки, да и то не мгновенной, а в случае неустранения выявленных нарушений доступа к сервису. То есть в документах показаны ориентиры, выделены важные моменты, которые разработчики не должны упускать из виду, но никакой системы наказаний за несоблюдение установленных норм нет. Словом, лейтмотив китайской регуляторики, как мне кажется, – не навредить развитию технологий ИИ слишком строгими ограничениями.

И четвертый путь – российский, он немного похож на американский, но с той лишь разницей, что в России большую роль играют саморегулируемые организации (СРО), в частности Альянс в области искусственного интеллекта, созданный представителями индустрии и академического сообщества. Все члены Альянса обязуются соблюдать определенные стандарты и практики в области разработки и применения систем ИИ. Это позволяет более гибко реагировать на изменения и вызовы, проблемы той же этики и пр. В принципе, потенциально вполне жизнеспособная модель, когда СРО, аккредитованные государством, утверждают стандарты и со временем могут взять на себя функцию выдачи соответствующих лицензий, если речь идет о каких-либо областях применения, требующих более строгого регулирования.

■ Но если у людей нет общего понимания того, что есть добро, справедливость, истина, то на основании какой «моральной карты» они будут «выравнивать» искусственный интеллект, чтобы он мог принимать этически взвешенные решения, согласованные с человеческими интересами? На мой взгляд, это одна из центральных проблем.

– К сожалению, многие предложения по регулированию этой области наивны в силу непонимания того, каким образом эти технологии работают. Давайте представим себе мир, в котором, например, создателям фотошопа сказали, что они не должны допускать возможность рисовать любой противозаконный или оскорбительный контент.

Появился бы в таком случае этот программный продукт? Нет. И каким бы неэтичным ни было изображение, его можно и карандашом нарисовать на бумаге, и краской на заборе. Значит ли это, что нужно ограничивать использование красок или карандашей? Так далеко можно зайти. Хотел бы обратить внимание на уже упомянутую мной угрозу, которую могут представлять сами регуляторы – опасность замедления развития технологического искусственного интеллекта. Если вернуться к вопросу о возможном апокалипсисе, который ждет человечество в связи с развитием ИИ, важно понимать, что мы в принципе живем в мире, где стремительно развиваются технологии. Так, если бы в XIX в. кто-то захотел нанести непоправимый ущерб всему нашему виду или, тем более, биосфере планеты, то у него для этого просто не было бы подходящих инструментов. Что он мог бы сделать? Собрать весь порох, который был в наличии, и взорвать его? Но это вряд ли привело бы к глобальной катастрофе. В середине прошлого века, с развитием ядерных технологий, мир стал ближе к апокалипсису. Но тогда была создана сложная система сдержек и противовесов, контроля, регламентов безопасности, снизившая вероятность опасного применения существующих технологий. В нынешнем веке помимо еще более развитых ядерных технологий и, соответственно, угроз, исходящих от них, добавились биотехнологии, обладающие немалым потенциалом разрушения. Но, как точно подметил персонаж супергеройского фильма про Человека-Паука, с большой силой должна приходиться большая ответственность.

Наша способность управлять технологической мощностью, которую мы формируем, растет не так быстро. Даже если механизмы отбора направлены на «поумнение» людей, в чем есть серьезные основания сомневаться, биологическая эволюция вряд ли может состязаться в скорости с научно-техническим прогрессом. Люди меняются мало. Да, разрыв может отчасти быть компенсирован развитием культуры и социальных институтов, но без соответствующих технических средств возможности такого развития тоже сильно ограничены. И такими техническими средствами как раз и являются технологии ИИ, цель которых – раздвинуть границы возможного для человеческого разума, выступить в роли инструментов, многократно усиливающих и расширяющих интеллектуальные способности человечества. Эти технологии в условиях существующих технологических угроз играют, что называется, на нашей стороне.

■ В производственной практике есть такое понятие «пока-йоке», что означает «защита от ошибок» или, проще говоря, «защита от человеческого фактора».

– Есть такая дисциплина – риск-менеджмент, и одним из базовых принципов, лежащих в основе анализа рисков, является исследование всех возможных сценариев. У нас, к сожалению, очень часто говорят о проблемах, связанных с выходом каких-то систем искусственного интеллекта из-под контроля, но никто не говорит о рисках, с которыми может столкнуться общество, если преднамеренно замедлить развитие ИИ. На мой взгляд, мы вполне можем совершить коллективную лоботомию и в результате уничтожить себя во множестве катастрофических сценариев просто потому, что будем интеллектуально обезоружены перед лицом тех или иных рисков. У человечества нет опции остановить научно-технический прогресс, и дело даже не в том, что это невозможно, это еще и крайне опасно – как резкое торможение на огромной скорости. Но я думаю, что в наших силах обеспечить сбалансированное устойчивое развитие, а оно невозможно без развития соответствующих технологий ИИ.

■ Вы упоминали архитектуру трансформера, на которой построены современные LLM. Это конечная ветвь на пути к общему ИИ, которую пришло время масштабировать, или существуют какие-то принципиально иные подходы?

– У трансформеров есть определенные ограничения, в том числе квадратичная алгоритмическая сложность полного блока самовнимания (self-attention), лежащего в основе трансформерных моделей. Фактически вычислительные затраты растут как квадратичная функция от длины анализируемого контекста. Эта проблема адресует немало интересных альтернатив и разных вариантов трансформеров, к примеру с разреженным полем внимания, или современные рекуррентные модели – RetNet, RWKV, Mamba и др. Пока что ни одна из них не показала радикального преимущества над трансформерами на широком круге задач, но в отдельных случаях они работают заметно лучше.

■ Может ли случиться так, что для решения разных задач будут привлекаться разные подходы?

– Думаю, да. И более того, на смену классическому трансформерному блоку может прийти какой-то его

обновленный аналог. Если сравнивать архитектуру блока трансформера в 2018 г. и то, что он представляет собой в современных моделях, то заметим существенные изменения. Довольно много проводится экспериментов по итеративному улучшению трансформера и по ускорению его работы. Поэтому я бы остерегся строить какие-то конкретные прогнозы и утверждать, что через 5 лет мы полностью уйдем от трансформеров и перейдем к рекуррентным архитектурам, хотя такое вполне может случиться.

■ В чем конкурентное преимущество российских разработок в этом ключе и насколько мощно они могут себя проявить?

– Наша традиционно сильная сторона – математическая школа. Технологии искусственного интеллекта неразрывно связаны с развитием математических моделей. Кроме того, мы обладаем передовым опытом создания больших инженерных проектов. И в этом смысле важно понимать, что в области ИИ во многом идет борьба за разработку не самых лучших архитектур, а производственных комплексов, которые сами в состоянии генерировать такие модели. А это включает в себя и организационную структуру, и технологическую инфраструктуру, и определенную культуру исследований. Мы стремимся построить «машину инноваций», которая позволит перейти от ставки на успехи талантливых одиночек к построению мощного научно-производственного конвейера, во много раз более эффективного, чем любая кустарщина.

■ То есть речь идет, например, о той же компании «Сбер»?

– Прежде всего о ней. И это можно отнести к третьей составляющей наших серьезных конкурентных преимуществ – развитость применения программной инженерии в реальном бизнесе. В этом смысле «Сбер» – один из лидеров информатизации банковских процессов, признанный во всем мире. У нас есть, по сути, большой работающий бизнес, который стремится внедрять инструменты ИИ. Он принимает, понимает, ставит цели и задачи и все время развивается. Есть достаточно мощная смычка науки и производственного сектора, в данном случае – финансового. И это тоже наш козырь. Понятно, что существуют определенные вызовы, которые нужно преодолевать. Сейчас таковыми во всем мире являются ограниченные вычислительные мощности, используемые для обучения больших моделей в условиях высокой степени монополизации, в частности в электронике.

Тем не менее нам удастся адаптироваться к текущим реалиям, находить решения. Проблема технологического неравенства существует, некоторые страны стремятся ограничить доступ конкурентов к передовым технологиям.

■ Так было всегда, тут удивляться нечему.

– Отчасти верно, но этот вызов особенно актуален сегодня в силу того, что западные страны ужесточают механизмы экспортного контроля. Но в целом есть причины для некоторого ограниченного оптимизма – классическая электроника находится на финальной стадии жизненного цикла своего развития, и не исключено, что через 10–20 лет ей на смену придут принципиально другие вычислительные устройства, которые могут изменить вычислительный ландшафт. И в условиях этих изменений, я считаю, Россия вполне может занять достойное место в области развития технологий искусственного интеллекта.

■ Гонка за параметрами и объемом данных для обучения моделей продолжается. Есть ли у этого процесса физические и экономические пределы? Не упрутся ли разработчики в стену, когда стоимость обучения одной модели будет сравнима с ВВП небольшой страны?

– Те тренды, которые любят рисовать в виде экспонент, часто лишены всякого смысла. Например, говорить о росте числа параметров моделей бессмысленно, если не брать в расчет время обучения моделей и количество обучающих данных. Сколько бы параметров ни было у модели, если вы будете обучать ее недолго или на крайне ограниченном наборе данных – результаты будут плачевными. Реальные законы масштабирования систем определяются физическими ограничениями и инженерными возможностями. Например, нельзя масштабировать энергетику теми же темпами, которыми сегодня масштабируются вычислительные бюджеты на обучение больших моделей. Современная электроника потребляет довольно много энергии. Но это не фундаментальное физическое ограничение, а инженерное. И оно, скорее всего, будет со временем преодолено. Например, активно ведутся исследования в области нейроморфных систем, энергопотребление которых может быть на порядки ниже. В конце концов, человеческий мозг потребляет всего 20 Вт энергии, обеспечивая достаточную производительность для решения самых сложных задач.

Есть и более серьезные барьеры, которые нельзя преодолеть без ревизии всей физики, чего вряд ли стоит ожидать. Например, лимит Бремермана (в современном виде представленный в виде теоремы Марголуса – Левитина), принцип Ландауэра, алгоритмическая сложность – важнейшие концепции, устанавливающие фундаментальные ограничения на обработку информации и связанную с ней работу. До бесконечности наращивать интеллектуальную мощь системы нельзя. Если говорить о нынешних вызовах, то один из них связан с тем, что темпы роста объемов оцифрованных данных медленнее, чем темпы роста вычислительных мощностей. Конечно, объемы данных растут, причем сверхлинейными темпами, даже если мы посмотрим на тексты. Однако в самом ближайшем будущем существует риск оказаться в ситуации, когда будут ограничения, связанные именно с объемами данных, поэтому нужны методы, позволяющие использовать имеющиеся данные более эффективно. Сегодня многие исследования ведутся именно в этой области. Пока что горизонтальное масштабирование за счет добавления новых данных и вычислительных мощностей – самый простой способ улучшения моделей. Но это вовсе не значит, что не анализируются альтернативы и не изучаются подходы, которые позволят в условиях ограничения по данным делать модели более и более умными.

Кстати, создание мультиагентных систем – один из таких подходов. Но в целом есть много перспективных направлений, позволяющих из уже имеющихся данных получить в условиях увеличения вычислительного бюджета более совершенные системы. Это и создание синтетических данных, переупорядочение обучающих наборов данных (curriculum learning – обучение на основе учебных курсов), переход от чисто декодерных трансформерных моделей к энкодер-декодерным, в том числе диффузионным, которые на моменте предобучения решают не задачу авторегрессионной генерации текстов, а задачу денойзинга – удаления специально добавленного в данные шума. Так что здесь никакого фундаментального перелома, скорее всего, не случится, но опять же, горизонтальное масштабирование тоже долго продолжаться не может: постепенно отдача от него будет сокращаться, а в целом развитие способностей больших моделей будет продолжаться. Важно понимать, что гипотеза о том, что нас ждет интеллектуальный взрыв, когда модели научатся совершенствовать сами себя и уйдут в отрыв, – сценарий, конечно,

нереализуемый по ряду вполне объективных причин. Это такая немножко ненаучная фантастика. Тем не менее модели будут становиться умнее, решать задачи лучше, чем им это удастся сегодня. Seriously расширится их применение в самых разных областях экономики. В этом смысле, возвращаясь к вопросу, не может ли нас ждать новая «зима» искусственного интеллекта, думаю, что этот сценарий невозможен потому, что технологии ИИ очень плотно вошли в нашу жизнь.

■ Назовите, пожалуйста, одну область – медицину, науку, искусство, которая радикально преобразится под влиянием ИИ в ближайшие 3–5 лет.

– Самая горячая точка – робототехника, выход систем искусственного интеллекта в реальный физический мир благодаря появлению VLA-моделей (Vision-Language-Action), объединяющих в пространстве внутренних представлений естественный язык, зрение и действия исполнительных устройств. С появлением таких моделей разработчики впервые увидели путь к созданию действительно универсальных роботов.

■ То есть нас ждет производство без людей?

– Не совсем так, люди будут ставить производственные задачи, наблюдать за работой систем, совершенствовать их, сопровождать, корректировать производственные планы и т.д. Они никуда не денутся. Точно так же, как появление циркуля или фотоаппарата не привело к исчезновению изобразительного искусства. Просто теперь человек будет вооружен новыми инструментами и сможет концентрироваться не на решении каких-то низкоуровневых задач, а подняться на более высокие уровни абстракции.

■ Какой из внедренных проектов «Сбера», на ваш взгляд, на все 100% уникален?

– Магистральный флагманский проект GigaChat. Понятно, что это продукт коллективного труда большого количества команд, и под его «капотом» спрятано множество разных моделей – например, для рисования изображений GigaChat использует модель Kandinsky Image. GigaChat – современная система, продвинутый сервис генеративного искусственного интеллекта, работающий на наших языковых моделях, обучающихся с нуля на собранных и подготовленных нами данных. Мы к этому стремились, и у нас это получилось. Конечно, мы не придумываем собственные языки программирования или фрейм-

ворки. Хотя даже на уровне последних довольно активно переписываем код под собственные нужды. Кроме GigaChat нами развивается также платформа GigaCode – специализированный инструмент для помощи в разработке программного обеспечения. Семейство моделей Symformer предназначено для работы с музыкой, GigaCheck – для детекции сгенерированного контента и т.д.

■ Последняя модель Kandinsky – российский аналог Midjourney и Symformer для генерации музыки и вокального пения – это полностью ваш продукт?

– История «Кандинского» началась в 2021 г., когда появилась информация о создании OpenAI модели DALL-E, способной генерировать изображения по произвольному текстовому описанию. Нашим ответом стала работа над собственной моделью, получившей название ruDALL-E Malevich. Вслед за этой крошечной моделью появилась вторая версия, которую мы назвали ruDALL-E Kandinsky. И дальше, соответственно, стали ее развивать уже под брендом Kandinsky. Сегодня в семействе Kandinsky также развиваются модели Kandinsky 3D (для генерации трехмерных моделей), Kandinsky Video (для генерации видео), Malvina – уникальная модель для редактирования изображений при помощи инструкций на естественном языке.

■ По-моему, такие программы уже существуют в мире? В чем новизна вашей?

– Конечно, мы не единственные, кто разрабатывает модели для инструктивного редактирования картинок. Но мы пионеры в создании моделей для редактирования изображений с сохранением пиксельной стабильности. ChatGPT, к примеру, нельзя заставить редактировать картинку, сохранив при этом лица людей без изменений. Сегодня Google активно развивает похожую на Malvina модель Nano Banana, доступную в Gemini, есть еще несколько довольно интересных проектов в этой области. Но мы все-таки были первыми и сегодня в числе лидеров по ряду метрик.

■ Не могу не спросить вас относительно кадрового обеспечения для столь масштабных задач. И правда ли, что высшее образование теряет свою актуальность?

– Оно утратит свои позиции и сильно обесценится, только если останется неизменным. К счастью, высшая школа стремится адаптироваться, пусть не так быстро, как хотелось бы, но этот

процесс идет. Что касается нашей команды, то при приеме нового сотрудника наличие высшего образования всегда в приоритете – не столько с позиции обладания определенными знаниями, которые он получил в стенах учебного заведения, сколько в качестве подтверждения его усидчивости, способности думать, анализировать, работать системно. Наличие высшего образования – показатель определенной зрелости, если хотите, культурного кода. Оно увеличивает, улучшает, усиливает мягкие навыки человека, его способность искать информацию, вести исследования самостоятельно. Наконец, обучение в вузе банально расширяет кругозор. А если это ведущий университет, то это своеобразный знак качества, подтверждающий, что его выпускник обладает необходимыми компетенциями. В любом случае перед образованием стоят большие вызовы, связанные с быстрым развитием технологий. Если 50 лет назад те знания, которые старшее поколение получало в стенах вузов, могли оставаться востребованными на производстве долгие годы, то сейчас фактически каждый специалист, работающий в области машинного обучения, да и в целом в IT-сфере, обязан постоянно совершенствоваться и осваивать новые подходы. Поэтому ему нужен, скорее, метанавык – умение учиться. Насколько современные учебные программы соответствуют развитию именно этих сильных сторон своих студентов, тут большой вопрос. Характерный пример – отношение педагогов к применению студентами сервисов генеративного ИИ.

Проблему нельзя решить одними только запретами и ограничениями – мы должны учить молодых людей грамотно пользоваться этими инструментами. Все это до боли похоже на историю моего поколения, когда в школе учительница математики говорила, что у нас в кармане не будет калькулятора. Как показала жизнь, теперь калькулятор всегда при нас.

Вы полагаете, что нужно готовить общество к миру, в котором технологии ИИ не только будут доступны, но и станут неотъемлемой частью нашей жизни?

– Опыт «Сбера» в этой части уникален, потому что в его структуре функционирует СберУниверситет, который занимается на систематической основе переподготовкой кадров. И ребята из нашей команды, и те, кто работает в бизнесе, помогают в создании адаптированных программ, непосредственно участвуют в процессе обучения. Этому

направлению придается очень большое значение. Важно, чтобы люди подтягивались под тот уровень развития технологий, который у нас есть.

И финальный вопрос: от чего вы хотели бы предостеречь общество?

– В эпоху стремительного прогресса технологий парадоксальным образом в первую очередь востребованы именно человеческие качества. Как мы применим все новшества, зависит исключительно от нас с вами. При помощи молотка можно забивать гвозди, строить дома, а можно проломить чью-то голову. Что будут люди делать с технологиями, зависит исключительно от них. Поэтому, мне кажется, чрезвычайно важно обратить внимание на то, каким образом мы применяем новые технологии, в каких целях, на чьей стороне они будут действовать – добра или зла. Очень хотелось бы, чтобы мощные современные разработки приносили обществу больше пользы, чем вреда. И это зависит от всего общества в целом. И вторая вещь, на которую хотелось бы обратить внимание: мир становится все более и более динамичным, и для того, чтобы занять в нем достойное место, важно задумываться о своей карьере, развитии тех качеств, которые сделают вашу жизнь интересной и насыщенной. Способность учиться, пробовать что-то новое, адаптироваться к быстро изменяющейся технологической среде – вызов текущего момента. Не бойтесь идти вперед, смелее экспериментируйте, запускайте собственные проекты, превращайте хобби в реальное дело. И не слушайте тех, кто говорит, что что-то невозможно. Именно смелость и открытость новому, которые сегодня есть у молодежи, должны стать основой нашего будущего. ■

Жанна КОМАРОВА