



Генеративный ИИ в образовании: проблемы и перспективы

Технологии искусственного интеллекта (ИИ) известны с середины XX в. и в последние три года вышли на уровень выполнения творческих функций, традиционно считавшихся прерогативой человека. Такой качественный скачок связан с усовершенствованием алгоритмов обработки естественного языка, вызвавших огромный интерес у представителей бизнес-сообществ, различных групп пользователей, а также органов государственного управления. На генеративный ИИ стали возлагать большие надежды, считая его прорывным инструментом как для экономики, так и для социальной сферы. Как отмечено в подготовленном под эгидой ООН Докладе о человеческом развитии 2021–2022, «увеличение того, чего люди могут достичь, используя ИИ в качестве дополнения, а не заменяя то, что люди могут делать, открывает огромные перспективы» [1].



Дмитрий Мазарчук,
проректор по научно-методической работе
Университета НАН Беларуси,
кандидат исторических наук,
доцент

Изучение этого феномена на практике еще продолжается. Однако уже сейчас ясно, что его применение серьезно корректирует многие сложившиеся подходы к организации производства, потребления, коммуникации, намечает грядущие изменения правовой, социальной и социокультурной нормативности. Дискуссии с участием ведущих компаний, учреждений образования и науки, международных и межправительственных организаций на эту тему ведутся повсеместно.

Одной из важнейших тем в обсуждениях являются этические вопросы его внедрения в различных сферах деятельности. Множество примеров побуждают специалистов и государственных регуляторов к осторожности. Все чаще звучат призывы к ограничению дальнейших работ по машинному обучению больших языковых моделей, в некоторых случаях находящие отражение на законодательном уровне. Например, 14 июня 2023 г. в Европейском союзе принят Акт об искусственном интеллекте, 15 августа 2023 г. в Китайской Народной Республике – Временные меры по управлению услугами генеративного искусственного интеллекта.

В нашей стране необходимость разработки и широкомасштабного внедрения современных технологий ИИ закреплена Государственной программой инновационного развития на 2021–2025 гг., а также другими нормативными правовыми актами законодательства и подзаконными актами Правительства и органов государственного управления.

Как отмечено в Концепции национальной безопасности Республики Беларусь, «информационная сфера превращается в системообразующий фактор жизни людей, обществ и государств» [2]. Искусственный интеллект играет решающую роль в переходе к новому технологическому укладу, иногда определяемому как Четвертая промышленная революция (Индустрия 4.0). Буквально каждый день открываются новые возможности его использования в различных сферах, в том числе и в образовании.

Генеративный искусственный интеллект представляет собой тип ИИ, способный создавать контент на основе больших языковых моделей (LLM), иначе называемых алгоритмами обработки естественного языка (NLP). Принцип их действия основан на подборе наиболее вероятного следующего токена (единицы учета – например, слова) в цепочке. Параметром подбора служит весовой



Рис. 1. Обзор больших языковых моделей

коэффициент токена. Таким образом, подстановка каждого следующего слова в тексте осуществляется исходя из вероятности взаимосвязей между лексемами.

Вывод на пользовательский рынок компанией OpenAI алгоритма GPT-3 (июнь 2020 г.), чат-бота ChatGPT (ноябрь 2022 г.) и алгоритма GPT-4 (март 2023 г.) интенсифицировал усилия других ведущих разработчиков в области искусственного интеллекта. На *рис. 1* видно, что в последние годы происходил экспоненциальный рост глобального рынка алгоритмов обработки естественного языка. На сегодняшний день текущий момент он оценивается в 18,9 млрд долл. с прогнозом более 68 млрд в 2028 г.

Обучение больших языковых моделей осуществляется прежде всего на масштабных корпусах созданных людьми текстов, включая взятые из Интернета. Большой объем контента и большее число заложенных параметров (весовых коэффициентов) обеспечивают лучший, с точки зрения человека, результат при генерировании. Это может быть имитация человеческого языка с такими его особенностями, как эмоции, юмор, контекст. Для примера: модель GPT-3 прошла обучение на 570 Гб текста (почти 500 млрд токенов) и учитывает 175 млрд параметров.

Появление доступных больших языковых моделей сопровождалось активным обменом опытом их использования и обсуждением последующего применения в разных сферах деятельности. Особый интерес вызвала функция обратной связи, имитирующей диалог с человеком. В настоящее время генеративный ИИ приобрел мультимодальный характер. Это значит, что его алгоритмы или модели

способны производить контент в разных модальностях – текст, изображения, программный код, аудио, видео – и сочетать их. Например, алгоритмы DALL-E, Imagen, Midjourney и им подобные способны преобразовывать текстовый запрос пользователя (промпт) в изображение. Плагины ChatGPT позволяют строить диаграммы и графики, рисовать карты, пересказывать видеодорожки и даже писать промпты для Midjourney.

Возможности генеративного ИИ широко используются в самых разных сферах, наиболее часто – в медицине, масс-медиа, рекламе, маркетинге, копирайтинге и развлечениях. Как отметил аналитик компании Constellation Research Дион Хинклифф, благодаря способности генеративного искусственного интеллекта создавать синтезированный текст, изображения и другие типы контента практически без усилий скорость выполнения многих операций увеличится в 1,3–5 раз [3].

В 2023 г. группой исследователей на основе объемной базы данных O*NET был проведен анализ профессий, рабочие задачи в которых могут быть существенно интенсифицированы с помощью чат-ботов. Авторы пришли к выводам, что подавляющее число сотрудников в будущем неизбежно должны почувствовать на себе влияние генеративного ИИ. Примерно 19% из них могут столкнуться с тем, что будут автоматизированы не менее 50% выполняемых ими функций, а для 80% – как минимум 10%. К числу уязвимых относятся прежде всего интеллектуальные профессии [4].

Распространение генеративного ИИ вызвало к жизни особые механизмы, улучшающие

результаты, расширяющие сферу его применения и требующие так называемой «тонкой настройки». Она представляет собой цепочку промптов, в результате использования которых решение поставленной задачи достигается ограниченным множеством параметров.

По данным компании Gartner, в настоящее время большие языковые модели находятся на пике чрезмерных ожиданий (*рис. 2*). На этом этапе они приобретают популярность и становятся предметом широкого обсуждения, от них ожидают революционных свойств. Следует учесть, что за пиком следует период избавления от иллюзий, свойственных предыдущему этапу [5]. Поэтому особенно важно выявить присутствие генеративному ИИ недостатки, ограничения и риски.

Принцип действия больших языковых моделей предполагает генерирование информации исходя из структурных связей между лексемами и заложенного в предыдущий текст смысла. Однако эти модели не обладают знанием о реальности, которую выражена в языке, визуальном или аудиоконтенте знаковая система всего лишь описывает.

Хотя модели производят впечатление творческих, генерируемый ими продукт нередко логически бессмыслен либо недостоверен. Эту ошибку называют «галлюцинацией» больших языковых моделей. Примерами являются выдуманные цитаты, ложные факты, ссылки на несуществующие источники или зафиксированная в некоторых случаях пристрастность (расовая, религиозная, политическая).

Минимизация вероятности этой ошибки достигается путем промпт-инжиниринга – последовательной серии правильно составленных запросов, а также

проверки полученных фактов, пользовательского ограничения степени креативности модели, дополнительных корректирующих промптов, использования обучения с подкреплением с обратной связью (RLHF).

Сдерживает применение больших языковых моделей ограничение длины контекста. К примеру, в популярной модели GPT-3.5 она составляет 16 тыс. токенов, в более современной GPT-4 Turbo – 128 тыс. (на русском языке – примерно 300 тыс. знаков). В результате в рамках одного сеанса алгоритм не может справиться с объемной пользовательской задачей либо буквально «забывает» начало цепочки промптов.

Решением данной проблемы является максимальная формализация запросов – разбиение сложного задания на ряд небольших подзадач (в частности, так рекомендуют делать запросы при составлении книг).

Зависимость генеративного ИИ от обучающего корпуса и связанный с ней недостаток контекстуального знания (прежде всего о произошедших с момента обучения модели событиях) приводят к неспособности самообучения. Алгоритмы ИИ не могут автоматически обновляться, что сужает пространство их применения, они не способны к интуиции и творческому мышлению, что ограничивает их способности к глубокому анализу. Сгенерированные правильные с точки зрения грамматики и семантических связей тексты порождены статистическими закономерностями, а не знанием природы вещей. Несмотря на разработанные уловки и промпты, большие языковые модели не в состоянии выполнять многомерные проекты, предусматривающие целостную картину и динамические связи.

Подобно любому другому механическому устройству или программному приложению, генеративный ИИ не имеет этической осознанности и действует в соответствии с заданными правилами и статистическими закономерностями, не обладая моральными убеждениями или сознательным выбором. В то же время генерируемый ими контент не является этически нейтральным, поскольку неизбежно демонстрирует существующие на данном этапе развития социума предубеждения, прямо или косвенно отраженные в обучающем корпусе и запросах пользователей, может носить дискриминирующий, ненавистнический, предвзятый, экстремистский характер.

Как правило, в условиях применения больших языковых моделей оговаривается, что «пользователь несет полную ответственность за контент» [6]. Однако особенность человеческого восприятия такова, что происходит субъективация алгоритмов ИИ, которые наделяются признаками мыслящего индивида, полноценного участника коммуникации. Поэтому генерирование контента по запросам пользователей потенциально ведет к опасности

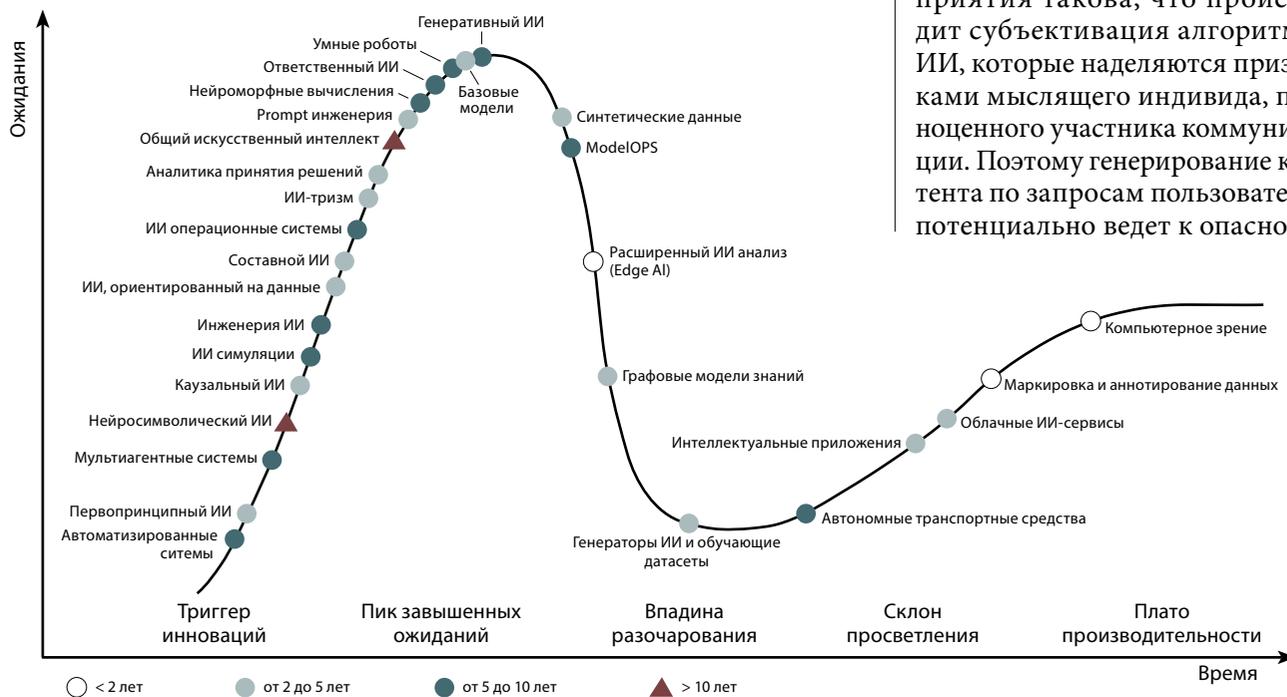


Рис. 2. ChatGPT и искусственный интеллект в высшем образовании. Краткое руководство ЮНЕСКО

распространения вводящей в заблуждение информации, неправомерного воздействия на общественное мнение, нарушения границ национальной безопасности государств. Добавим, что архитектура больших языковых моделей содержит в себе потенциальную угрозу неприкосновенности частной жизни и больших данных, а также размывает понятие интеллектуальной собственности [7].

Вторая группа проблем этического свойства связана с неравномерным уровнем интеграции информационно-коммуникативных технологий в производство и потребление, недостаточным доступом некоторых регионов земного шара к Интернету. Более того, заложенный в механизмах ИИ экономический потенциал может привести к существенным структурным изменениям в характере работы и труда, усугубляя существующее глобальное неравенство.

Большие языковые модели представляют собой мощный инструмент научного познания и организации труда ученого, оптимизируя многие исследовательские задачи путем анализа больших данных, майнинга научных публикаций, машинного перевода и коммуникаций между учеными разных стран, расширяя границы научного прогресса [8]. Уже сейчас признаются большие возможности использования генеративного ИИ в молекулярных и геномных изысканиях, при моделировании сложных процессов (погодные условия, биоразнообразие), оптимизации энергетических систем. Однако резкое ускорение инноваций и научных открытий может привести к снижению стандартов безопасности и этики, а значит, еще больше повысить социальные риски.

В наше время проблема подконтрольности генеративного ИИ человеку из области фантастики переходит в практическую плоскость. На начальном этапе его развития вопросы этики обсуждались экспертами и пользовательским сообществом. Первые этические кодексы в данной сфере разработаны ведущими компаниями (Microsoft, IBM) и авторитетными международными организациями. Ключевыми документами являются Пекинский консенсус по искусственному интеллекту и образованию (ЮНЕСКО, 2019 г.), Рекомендация об этических аспектах искусственного интеллекта (ЮНЕСКО, 2021 г.), Промежуточный отчет Консультативного органа по искусственному интеллекту при Генеральном секретаре ООН (декабрь 2023 г.). В марте 2023 г. несколько экспертов и менеджеров крупных технологических компаний подписали призыв к ограничению разработок генеративного ИИ [9]. В последующем обсуждение этого вопроса вышло на уровень правительств отдельных государств и международных организаций. Работа по созданию всеобъемлющего механизма регулирования ИИ и рекомендаций по его применению продолжается.

Активное проникновение технологий генеративного ИИ в сферу образования началось с релиза чата ChatGPT; впоследствии стало очевидно, что потенциал ИИ в этом сегменте огромен и в значительной степени пока что не использован. Его применение может улучшить персонализированное обучение, обеспечить эффективное управление, расширить исследовательский потенциал и облегчить международное сотрудничество, помочь учебным заведениям в выпол-

нении их социальных функций и тем самым содействовать достижению Целей в области устойчивого развития (ЦУР).

Множество примеров демонстрирует поистине впечатляющие успехи больших языковых моделей. Так, на официальном сайте модели GPT-4 представлены следующие достижения: 89-й перцентиль в академическом тесте по математике, 90-й перцентиль на экзамене на адвокатскую практику, 99-й перцентиль на олимпиаде по биологии. Хорошо известен случай защиты выпускником одного из российских вузов дипломного исследования, подготовленного им путем серии системных промптов в ChatGPT. Затраченное время на работу, получившую высокую оценку (с 82% оригинальностью), составило 38 часов. Это позволяет ставить вопрос о пересмотре традиционных методов оценивания в образовании.

Благодаря своим способностям генерировать и оценивать информацию генеративный ИИ может выполнять ряд функций в обучении как в качестве отдельных инструментов, так и в интеграции с другими платформами (таблица).

Недавний обзор 118 публикаций на тему использования больших языковых моделей для автоматизации образовательных задач выявил как ограничения, так и возможности в данной области [10]. Авторами было выделено 9 категорий таких задач, уровень автоматизации которых оценен как достаточно высокий. Например, созданные с помощью алгоритма BERT текстовые резюме не имели существенных отличий по сравнению с текстами, написанными студентами. Поскольку в большинстве проанализированных

в обзоре публикаций не раскрыты в достаточной степени применяемые методы исследований, этот опыт невоспроизводим, что является серьезным препятствием для внедрения образовательных инноваций, в связи с чем было рекомендовано повысить уровень их прозрачности с использованием генеративного ИИ (открытый код). К тому же более $\frac{3}{4}$ изученных публикаций находятся на стадии прикладных изысканий. Лишь в 7 из них описаны проверочные исследования использования LLM в аутентичной учебной среде.

Таким образом, инновации на основе генеративного ИИ еще не продемонстрировали доказательств улучшения преподавания, обучения и административных процессов в образовательной практике, поэтому «существует настоятельная потребность в реальных исследованиях, которые доведут основанные на больших языковых моделях инновации непосредственно заинтересованным сторонам в сфере образования» [11]. Они могут опровергнуть либо подтвердить высокую производительность моделей при выполнении задач

прогнозирования и генерации. Кроме того, особенно важным представляется предотвращение ненадлежащего использования таких технологий, что может осуществляться только на основе репрезентативной практики.

Авторы обзора подчеркнули большой потенциал генеративного ИИ в сфере образования. Очевидно, что на его раскрытие в ближайшем будущем будут оказывать влияние как обозначенные выше технологические и этические ограничения, так и существующие тренды развития искусственного интеллекта.

| Функция | Описание | Пример реализации |
|---|---|---|
| Механизм возможностей | ИИ генерирует альтернативные способы выражения идеи | Студенты пишут запросы в ChatGPT и используют функцию «сгенерировать ответ» для изучения альтернативных ответов |
| Диалоговая система с сократовой логикой | ИИ выступает в роли оппонента для развития навыков аргументации | Студенты вводят в ChatGPT запросы согласно структуре беседы или дебатов. Преподаватели могут попросить студентов использовать ChatGPT для подготовки к дискуссии |
| Инструмент сотрудничества | ИИ помогает группам студентов решать проблемы вместе | Работая в группах, учащиеся используют ChatGPT для поиска информации, необходимой для выполнения задач и заданий |
| «Проводник к знаниям» | ИИ выступает в роли проводника в изучении реальных и концептуальных тем | Преподаватели используют ChatGPT для создания материалов для занятий/курсов (например, вопросов для обсуждения) и получения советов о том, как помочь студентам в изучении конкретных концепций |
| Личный наставник | ИИ выступает в роли наставника для студента и дает немедленную обратную связь о ходе обучения | ChatGPT предоставляет студентам персонализированные отзывы на основе информации, предоставленной самими студентами или преподавателями (например, результатов тестов) |
| Помощник в разработке | ИИ оказывает помощь на протяжении всего процесса разработки | Преподаватели обращаются к ChatGPT за идеями для разработки или обновления учебной программы (например, процедуры оценивания) и/или для достижения конкретных целей (например, повышение доступности учебной программы) |
| Исследовательский центр | ИИ предоставляет инструменты для работы с данными, их изучения и интерпретации | Преподаватели предоставляют базовую информацию студентам, которые пишут различные запросы в ChatGPT, чтобы узнать больше по теме. ChatGPT можно использовать как помощника при изучении языка |
| Помощник в учебе | ИИ помогает студентам усвоить учебный материал | Студенты объясняют ChatGPT свой текущий уровень понимания материала и просят помочь им в изучении материала. ChatGPT также можно использовать, чтобы помочь студентам подготовиться к другим заданиям (например, к собеседованию) |
| Средство мотивации | ИИ предлагает игры и задачи, чтобы обогатить учебный процесс | Преподаватели и студенты обращаются к ChatGPT за идеями о том, как обогатить учебный процесс после предоставления краткой информации о текущем уровне знаний (например, викторины, упражнения) |
| Инструмент оценки знаний | ИИ предоставляет преподавателям срез текущих знаний каждого студента | Студенты взаимодействуют с ChatGPT в рамках диалога типа самостоятельной работы, а затем просят ChatGPT подготовить сводку по текущему уровню знаний, чтобы поделиться им с преподавателем или с целью оценки знаний |

Таблица. Возможности применения искусственного интеллекта в высшем образовании

Источник: ChatGPT и искусственный интеллект в высшем образовании. Краткое руководство. ЮНЕСКО // https://unesdoc.unesco.org/ark:/48223/pf0000385146_rus.

В 2024 г., по данным аналитической компании Forrester Research, они будут направлены на улучшение клиентского опыта и дальнейший рост персонализированных ИИ-решений [12]. В ближайшее время увеличится количество специализированных приложений (маркетинговых, логистических, медицинских, юридических и т.д.), что для индивидуальных пользователей будет означать повышение качества услуг. С другой стороны, для корпоративных клиентов открываются огромные возможности для анализа в маркетинговых целях больших данных.

Применительно к сфере образования ИИ выступает не только технологическим инструментом, но и средством для создания ее более эффективной, справедливой и инклюзивной модели. В первую очередь она предполагает автоматизацию множества ресурсоемких задач, в частности генерирования основных и иллюстративных учебных материалов, тестов и других средств контроля знаний, а также

системы оценивания результативности обучения на основе классификации когнитивных уровней.

Во-вторых, это персонализация обучения, то есть формирование с помощью моделей ИИ уникальных, индивидуализированных учебных модулей, адаптированных под потребности, особенности и возможности каждого обучающегося. Тем самым будут учитываться его психолого-педагогические характеристики; задаваться персональный темп при освоении отдельных дисциплин, модулей и тем; генерироваться персонализированные межпредметные связи, практические кейсы и задачи.

В-третьих, это качественный переход от взаимодействия на уровне «учитель – ученик» к полноценной интерактивной коммуникации с виртуальным тьютором.

В наши дни информационная среда превращается в системообразующий фактор жизни общества, включая сферу образования. Сотрудничество человека и ИИ

может стать важным навыком для обучающихся разных уровней (школа, вуз, аспирантура), нацеленных на успех в будущей профессиональной карьере, поскольку решения на основе технологий искусственного интеллекта становятся неотъемлемым компонентом современной экономики. Безусловно, внедрять большие языковые модели в образование следует взвешенно и осторожно. Логика развития Индустрии 4.0 ведет к необходимости совершенствования политики в отношении инструментов ИИ на основе эмпирических данных и отзывов заинтересованных сторон.

Не должна оставаться в стороне от этих процессов и Республика Беларусь. Максимальное использование возможностей генеративного ИИ при этически обоснованном и эффективном его регулировании – залог устойчивого развития и сохранения национального суверенитета в условиях жесточайшей международной экономической конкуренции. ■

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Доклад о человеческом развитии за 2021–2022 гг. «Времена неопределенности, неустойчивые жизни: наше будущее в меняющемся мире» // <https://hdr.undp.org/system/files/documents/global-report-document/hdr2021-22rupdf.pdf>.
2. Концепция национальной безопасности Республики Беларусь, утвержденная Указом Президента Респ. Беларусь от 09.10.2010 г. № 575 // <https://pravo.by/document/?guid=3871&p0=P31000575>.
3. Hinchcliff D. How Generative AI Has Supercharged the Future of Work // <https://www.constellationn.com/blog-news/how-generative-ai-has-supercharged-future-work>.
4. Eloundou T., Manning S., Mishkin P., Rock D. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models // <https://pravo.by/document/?guid=3871&p0=P31000575>. – DOI: 10.48550/arXiv.2303.10130.
5. What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle // <https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle>.
6. Open AI. Europe Terms of Use // <https://openai.com/policies/terms-of-use>.
7. Haoran Li [et al.]. Privacy in Large Language Models: Attacks, Defenses and Future Directions // <https://arxiv.org/pdf/2310.10383.pdf>. – DOI: 10.48550/arXiv.2310.10383.
8. AI Advisory Body. Interim Report: Governing AI for Humanity // <http://www.un.org/en/ai-advisory-body>. – Box 2.
9. Pause Giant AI Experiments: An Open Letter // <https://futureoflife.org/open-letter/pause-giant-ai-experiments>.
10. Lixiang Yan [et al.]. Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review // *British Journal of Educational Technology*. 2024. Vol. 55. P. 104.
11. Predictions 2024: Exploration Generates Progress // <https://go.forrester.com/wp-content/uploads/2023/10/Forrester-Predictions-2024.pdf>.