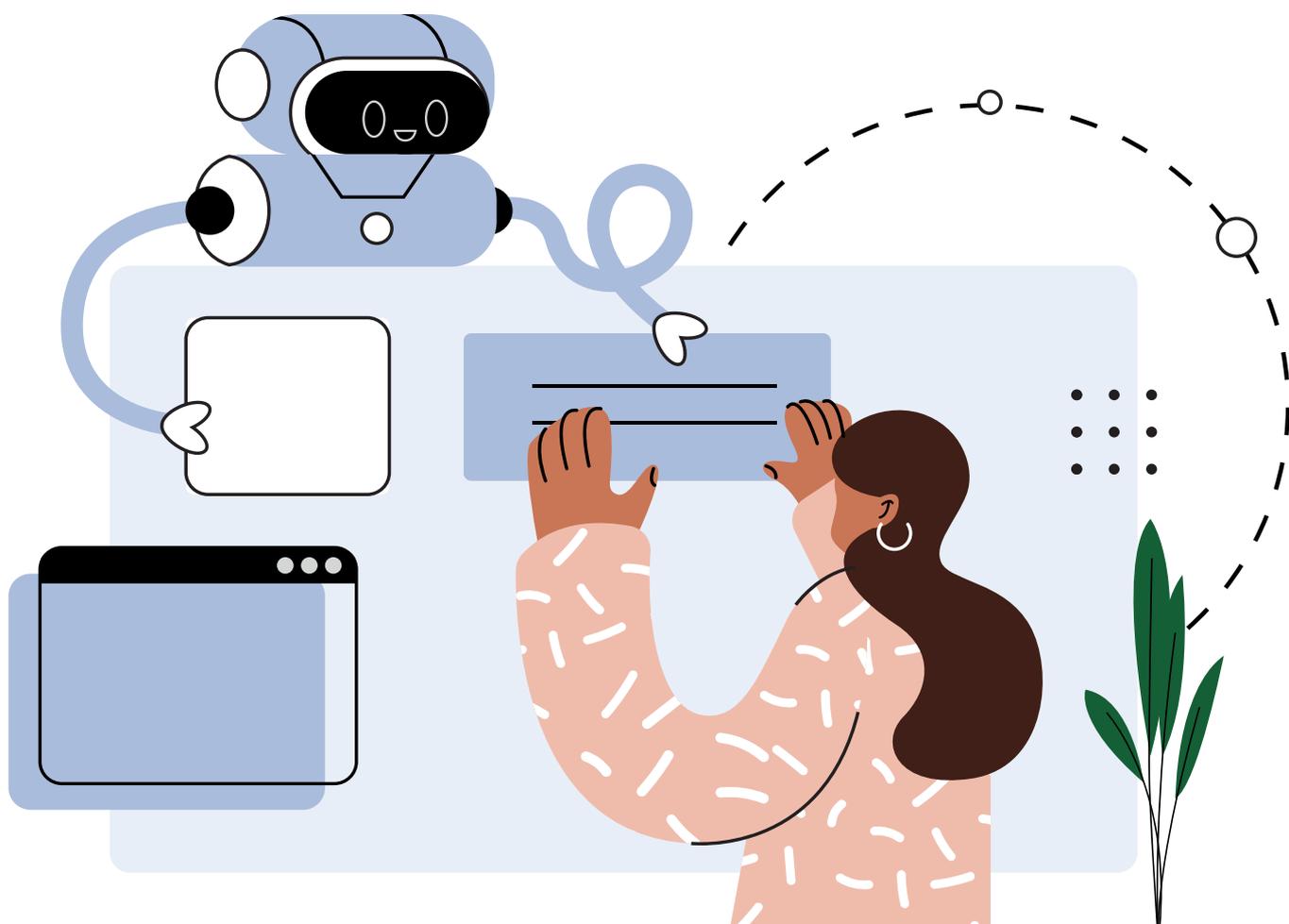




unesco

Руководство по использованию генеративного искусственного интеллекта в образовании и научных исследованиях



ЮНЕСКО – мировой лидер в области образования

Образование – высший приоритет ЮНЕСКО и является одним из основных прав человека и фундаментом мира и устойчивого развития. ЮНЕСКО – это специализированное учреждение Организации Объединенных Наций в области образования, обеспечивающее глобальное и региональное лидерство для достижения прогресса, повышения устойчивости и потенциала национальных систем для всех учащихся. ЮНЕСКО также прилагает усилия, чтобы ответить на современные глобальные вызовы посредством трансформирующего обучения с особым упором на гендерное равенство и Африку по всем действиям.

Глобальная повестка дня в области образования на период до 2030 года

ЮНЕСКО как специализированному учреждению Организации Объединенных Наций в области образования было поручено руководство и координация деятельности, связанной с осуществлением повестки дня «Образование-2030», которая является частью глобальных усилий по искоренению к 2030 году нищеты посредством реализации 17 целей в области устойчивого развития. Образованию, имеющему важнейшее значение для достижения всех этих целей, посвящена отдельная цель – Цель 4, направленная на «**обеспечение всеохватного и справедливого качественного образования и поощрение возможности обучения на протяжении всей жизни для всех**». В рамочной программе действий «Образование-2030» сформулированы руководящие принципы в отношении осуществления этой амбициозной цели и обязательства государств в этой области.



Опубликовано в 2024 г. Организацией Объединенных Наций по вопросам образования, науки и культуры
7, Place de Fontenoy, 75352 Paris 07 SP, France

© ЮНЕСКО 2024

ISBN 978-92-3-400077-2



Данная публикация предлагается в открытом доступе под лицензией Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). Используя содержание данной публикации, пользователи соглашаются с правилами пользования Репозитория открытого доступа ЮНЕСКО (<https://www.unesco.org/ru/open-access/cc-sa>).

Изображения, отмеченные звездочкой (*), не подпадают под действие лицензии CC-BY-SA и не могут быть использованы или воспроизведены без предварительного разрешения правообладателей.

Название оригинала: *Guidance for generative AI in education and research*

Опубликовано в 2023 году Организацией Объединенных Наций по вопросам образования, науки и культуры.

Использованные названия и представление материалов в данной публикации не являются выражением со стороны ЮНЕСКО какого-либо мнения относительно правового статуса какой-либо страны, территории, города или района или их соответствующих органов управления, равно как и линий разграничения или границ.

Ответственность за взгляды и мнения, высказанные в данной публикации, несут авторы. Их точка зрения может не совпадать с официальной позицией ЮНЕСКО и не накладывает на Организацию никаких обязательств.

Перевод на русский язык был подготовлен при поддержке Института ЮНЕСКО по информационным технологиям в образовании (ИИТО ЮНЕСКО).

Перевод: Анастасия Ляпичева

Редакция перевода: Галина Удайадас (ИИТО ЮНЕСКО)

Иллюстрация для обложки: Olexandra Simkina/Shutterstock.com*

Оформлено и подготовлено к печати ЮНЕСКО

Отпечатано во Франции

Применение ориентированного на человека подхода к использованию генеративного искусственного интеллекта

Инструменты генеративного искусственного интеллекта (генеративного ИИ), доступные для общего использования, развиваются быстрыми темпами, и выпуск итеративных версий опережает адаптацию национального законодательства. Отсутствие национальных норм относительно генеративного ИИ в большинстве стран оставляет конфиденциальность данных пользователей незащищенной, в то время как образовательные учреждения в значительной степени не готовы к проверке данных инструментов.

Первое глобальное руководство ЮНЕСКО по использованию генеративного искусственного интеллекта в образовании направлено на поддержку стран в проведении немедленных действий, разработке долгосрочной политики и развитии человеческого потенциала для реализации ориентированного на человека подхода к использованию этих новых технологий.

В руководстве представлена оценка потенциальных рисков, связанных с применением генеративного ИИ для важных гуманистических ценностей: участия в общественной жизни, инклюзивности, справедливости, гендерного равенства, языкового и культурного многообразия, а также свободы выражения мнений.

Документ включает ключевые шаги для регулирования использования инструментов генеративного ИИ со стороны правительственных органов, включая обязательную защиту конфиденциальности данных и рассмотрение возрастных ограничений для их применения. Также в руководстве представлены требования к поставщикам генеративного ИИ, направленные на обеспечение их этичного и эффективного использования в образовании.

В руководстве подчеркивается необходимость подтверждения образовательными учреждениями этичности и педагогической целесообразности использования систем генеративного ИИ в образовании. Документ призывает международное сообщество задуматься над долгосрочными последствиями для процесса усвоения знаний, преподавания, обучения и оценки.

Публикация предлагает конкретные рекомендации для политиков и образовательных учреждений относительно того, как развитие использования инструментов генеративного ИИ может способствовать защите человеческой деятельности и приносить реальную пользу студентам, учащимся и исследователям.

Несмотря на то, что в январе 2023 года ежемесячное количество пользователей **ChatGPT** достигло

100 миллионов

человек, только **одна страна** ввела регулирование использования **генеративного ИИ** в июле.



unesco

«Мысли о войне возникают в умах людей, поэтому в сознании людей следует укоренять идею защиты мира»



unesco

Руководство по использованию
генеративного искусственного
интеллекта в образовании и
научных исследованиях

Предисловие



© UNESCO

Генеративный искусственный интеллект (ИИ) привлек внимание общественности в конце 2022 года с запуском ChatGPT, который стал самым быстрорастущим приложением в истории. Эти приложения генеративного ИИ, способные имитировать возможности человека по созданию таких материалов, как тексты, изображения, видео, музыка и программные коды, вызвали бурную реакцию общества. Сегодня миллионы людей пользуются инструментами генеративного ИИ в повседневной жизни, и потенциал адаптации моделей к специфичным для конкретной области приложениям ИИ кажется безграничным.

Такие широкие возможности обработки информации и производства знаний потенциально имеют огромные последствия для образования, поскольку они воспроизводят мышление более высокого порядка, которое составляет основу человеческого обучения. Поскольку инструменты генеративного ИИ все больше способны автоматизировать некоторые базовые уровни письма и создания произведений искусства, они вынуждают политиков и

образовательные учреждения переосмыслить, почему, чему и как мы учимся. На данном этапе это критически важные соображения для образования на новом этапе цифровой эры.

Настоящий документ разработан с целью поддержки разработки соответствующих нормативных положений, политики и развития человеческого потенциала с целью обеспечения того, что генеративный ИИ используется в качестве инструмента, приносящего реальную пользу и расширяющего возможности для учителей, учащихся и исследователей.

Руководство содержит важные шаги для государственных учреждений по регулированию использования генеративного ИИ. В нем представлены основы и конкретные примеры для формулирования политики и разработки учебных программ, которые обеспечивают этическое и эффективное использование данной технологии в образовании. Наконец, документ призывает международное сообщество рассмотреть глубокие долгосрочные последствия использования генеративного ИИ для того, как мы понимаем, структурируем и оцениваем знания, а также оцениваем и подтверждаем учебные процессы.

Основанное на «Рекомендации ЮНЕСКО об этических аспектах искусственного интеллекта» 2021 года, данное руководство отражает гуманистический подход к образованию, способствующий развитию человеческой деятельности, инклюзивности, справедливости, гендерного равенства, а также культурного и языкового разнообразия, включая плюрализм мнений и выражений. Кроме того, данная публикация развивает идеи, представленные в докладе Международной комиссии по будущему образования 2021 года «Переосмыслим наше будущее – Новый общественный договор в области образования», который направлен на пересмотр наших отношений с технологиями как неотъемлемой частью усилий по обновлению общественного договора в сфере образования.

Искусственный интеллект не должен лишать человека его роли в интеллектуальной сфере. Напротив, он предлагает нам пересмотреть устоявшиеся представления о человеческих знаниях и обучении. Мы надеемся, что это руководство поможет нам открыть новые перспективы в образовании и послужит основой для коллективного мышления и объединенных действий, способных привести к созданию будущей цифровой образовательной среды, ориентированной на человека и доступной каждому.

Стефания Джаннини,
Заместитель Генерального директора ЮНЕСКО
по вопросам образования

Благодарности

Под руководством Стефании Джаннини, заместителя Генерального директора ЮНЕСКО по вопросам образования, и Собхи Тавил, директора Отдела будущего обучения и инноваций ЮНЕСКО, подготовку публикации возглавил Фэнчунь Мяо, руководитель Отдела технологий и искусственного интеллекта в образовании.

Особая благодарность выражается Уэйну Холмсу, доценту Университетского колледжа Лондона, который был соавтором некоторых разделов публикации.

Эта публикация является плодом коллективных усилий лидеров образования и экспертов в области искусственного интеллекта и образования.

В ней были использованы идеи и вклад многих экспертов, в том числе: Мутлу Чукурова, профессора Университетского колледжа Лондона; Колина де ла Игера с кафедры ЮНЕСКО по технологиям подготовки учителей с использованием открытых образовательных ресурсов в Нантском университете; Шафики Айзекс, научного сотрудника Йоханнесбургского университета; Натали Лао, исполнительного директора App Inventor Foundation; Цинь Ни, доцента Шанхайского педагогического университета; Каталины Николин, эксперта по ИКТ в образовании Европейского центра цифрового образования в Румынии; Джона Шоу-Тейлора с кафедры ЮНЕСКО по искусственному интеллекту и профессора вычислительной статистики и машинного обучения в Университетском колледже Лондона; Келли Широхира, исполнительного директора Jet Education Services; Ки-Санга Сона, профессора Корейского национального педагогического университета; Илкка Туоми, главного научного сотрудника компании Meaning Processing Ltd в Финляндии.

Многие коллеги из ЮНЕСКО также внесли свой вклад, в том числе: Дафна Фейнхольц, руководитель отдела биоэтики и этики науки и технологий; Франческо Педро, директор Международного института ЮНЕСКО по высшему образованию в Латинской Америке и Карибском бассейне; Пратика Сибала, программный специалист отдела цифровой политики и цифровой трансформации; Саураб Рой, старший специалист по проектам отдела подготовки учителей, отдела политики и систем непрерывного обучения; Бенджамин Вергель Де Диоса, программный специалист по ИКТ в области образования из отдела образовательных инноваций и развития навыков в офисе в Бангкоке; коллеги из Управления по разнообразию форм культурного самовыражения в секторе культуры; Марка Уэста, программного специалиста отдела будущего обучения и инноваций.

Благодарность также выражается Глену Хертеленди, Луизе Ферраре и Сянлей Чжэну из Отдела технологий и искусственного интеллекта в образовании, за координацию выпуска публикации.

Благодарность также выражается Дженни Вебстер за редактирование и корректуру текста, а также Нгок-Туи Тран за разработку макета.

Содержание

Предисловие	2
Благодарности	3
Список аббревиатур и сокращений	6
Введение	7
1. Что такое генеративный ИИ и как он работает?	9
1.1 Что такое генеративный ИИ?	9
1.2 Как работает генеративный ИИ?	9
1.2.1 Как работают текстовые модели генеративного ИИ	10
1.2.2 Как работают модели изображений генеративного ИИ	13
1.3 Оптимизация запросов для достижения желаемых результатов	14
1.4. Новый EdGPT и его последствия	15
2. Разногласия по поводу генеративного ИИ и их значение для образования	17
2.1 Усугубление цифровой бедности	17
2.2 Опережение адаптации национального регулирования	17
2.3 Использование контента без согласия владельцев данных	18
2.4 Особенности «необъяснимых» моделей в процессе генерации результатов	18
2.5 Искусственный интеллект и загрязнение контента в сети	19
2.6 Недостаточное понимание реального мира	20
2.7 Ограничение разнообразия мнений и дальнейшая маргинализация голосов	21
2.8 Создание более реалистичных дипфейков	21
3. Регулирование использования генеративного ИИ в образовании	22
3.1 Ориентированный на человека подход к искусственному интеллекту	22
3.2 Шаги по регулированию генеративного ИИ в образовании	22
3.3 Положение о генеративном ИИ: ключевые элементы	25
3.3.1 Государственные регулирующие органы	25
3.3.2 Поставщики инструментов генеративного ИИ	27
3.3.3 Институциональные пользователи	28
3.3.4 Индивидуальные пользователи	29
4. На пути к созданию политической основы для использования генеративного ИИ в образовании и исследованиях	30
4.1 Содействие инклюзивности, равенству, языковому и культурному разнообразию	30
4.2 Защита человеческой свободы действий	31
4.3 Мониторинг и проверка систем генеративного ИИ для образования	32
4.4 Развитие у учащихся компетенций в области ИИ, включая навыки, связанные с генеративным ИИ	32

4.5	Наращивание потенциала преподавателей и исследователей для правильного использования генеративного ИИ	33
4.6	Продвижение плюрализма мнений и множественного выражения идей	33
4.7	Тестирование локально значимых прикладных моделей и формирование базы доказательств	34
4.8	Анализ долгосрочных последствий на межсекторальной и междисциплинарной основе	35
5.	Содействие творческому использованию генеративного ИИ в образовании и исследованиях	36
5.1	Институциональные стратегии, способствующие ответственному и творческому использованию генеративного ИИ	36
5.2	Подход, ориентированный на человека и педагогически целесообразное взаимодействие	37
5.3	Совместная разработка использования генеративного ИИ в образовании и исследованиях	37
5.3.1	Генеративный ИИ для исследований	38
5.3.2	Генеративный ИИ для поддержки обучения	39
5.3.3	Генеративный ИИ в качестве индивидуального наставника для самостоятельного освоения базовых навыков	41
5.3.4	Генеративный ИИ для содействия исследованиям или проектному обучению	43
5.3.5	Генеративный ИИ для поддержки учащихся с особыми потребностями	44
6.	Генеративный ИИ и будущее образования и исследований	47
6.1	Неисследованные этические проблемы	47
6.2	Авторское право и интеллектуальная собственность	47
6.3	Источники контента и обучения	48
6.4	Однородные ответы против разнообразных и творческих результатов	48
6.5	Переосмысление оценивания и результатов обучения	48
6.6	Мыслительные процессы	49
	Заключительные комментарии	50
	Список литературы	51
Список таблиц		
Таблица 1.	Методы, используемые в генеративном искусственном интеллекте	10
Таблица 2.	GPT от OpenAI	11
Таблица 3.	Совместная разработка использования генеративного ИИ в научных исследованиях	38
Таблица 4.	Совместная разработка использования генеративного ИИ для поддержки учителей и обучения	40
Таблица 5.	Совместная разработка использования генеративного ИИ в качестве индивидуального наставника для самостоятельного освоения базовых навыков в языках и искусстве	41
Таблица 6.	Совместная разработка использования генеративного ИИ для содействия исследованиям или проектному обучению	43
Таблица 7.	Совместная разработка использования генеративного ИИ для поддержки учащихся с особыми потребностями	45

Список аббревиатур и сокращений

Концепции и технологии

AGI	Искусственный общий интеллект
AI	Искусственный интеллект (ИИ)
ANN	Искусственная нейронная сеть
API	Интерфейс прикладного программирования
DAI	Распределенный искусственный интеллект
GAN	Генеративно-сопоставительные сети
GB	Гигабайты
GDPR	Общий регламент по защите данных Европейского Союза
GenAI	Генеративный искусственный интеллект
GPT	Генеративный предварительно обученный трансформер
ICT	Информационно-коммуникационные технологии (ИКТ)
LaMDA	Языковая модель для приложений диалога
LLM	Большая языковая модель
ML	Машинное обучение
VAE	Вариационные автокодировщики

Организации

AGCC	Правительственный облачный кластер искусственного интеллекта (Сингапур)
CAC	Управление киберпространством Китая
EU	Европейский союз (ЕС)
OECD	Организация экономического сотрудничества и развития (ОЭСР)
UNCTAD	Конференция ООН по торговле и развитию (ЮНКТАД)
UNESCO	Организация Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО)

Введение

Выход ChatGPT в конце 2022 года, первого простого в использовании инструмента генеративного искусственного интеллекта (ИИ), ставшего широкодоступным для общественности¹, за которым последовали итеративно более сложные версии, вызвал шок по всему миру и подогревает гонку среди крупных технологических компаний за позиционирование себя в области разработки моделей генеративного ИИ².

По всему миру, первоначальная обеспокоенность в образовательной сфере заключалась в том, что ChatGPT и аналогичные инструменты генеративного ИИ могут использоваться учащимися для обмана при выполнении заданий, тем самым подрывая ценность оценки обучения, сертификации и квалификации (Андерс, 2023). В то время как некоторые образовательные учреждения запретили использование ChatGPT, другие осторожно приветствовали появление генеративного ИИ (Тлили, 2023). Многие школы и университеты, например, приняли прогрессивный подход, полагая, что «вместо того, чтобы пытаться запретить использование инструментов генеративного ИИ, необходимо поддерживать учащихся и сотрудников в их эффективном, этичном и прозрачном использовании» (Рассел Групп, 2023). Этот подход признает, что широкодоступный генеративный ИИ, вероятно, будет только становиться более сложным и иметь как специфический отрицательный, так и уникальный положительный потенциал для образования.

Действительно, у генеративного ИИ есть множество возможных применений. Он может автоматизировать обработку информации и представление результатов во всех ключевых символических представлениях человеческого мышления. Это позволяет достигать конечных результатов путем предоставления первоначальных, несформированных знаний или результатов обучения, которые могут быть усвоены или доработаны в процессе обучения.

Освободив людей от некоторых категорий навыков мышления низшего порядка, это новое поколение инструментов ИИ может иметь глубокие последствия для того, как мы понимаем человеческий интеллект и обучение.

Но генеративный ИИ также вызывает множество неотложных проблем, связанных с такими вопросами, как безопасность, конфиденциальность данных, авторские права и манипулирование. Некоторые из них представляют собой более широкие риски, связанные с искусственным интеллектом, которые еще больше усугубились из-за генеративного ИИ, в то время как другие возникли недавно с появлением инструментов последнего поколения. На данном этапе крайне важно полностью понять и решить каждую из этих проблем.

Настоящее руководство призвано удовлетворить эту насущную потребность. Однако тематический набор руководств по использованию генеративного ИИ для образования не следует понимать как утверждение, что генеративный ИИ является решением фундаментальных проблем образования. Несмотря на активное упоминание генеративного ИИ в СМИ, маловероятно, что он в одиночку решит проблемы, стоящие перед системами образования во всем мире. Реагируя на давние проблемы образования, важно поддерживать идею о том, что человеческий потенциал и коллективные действия, а не технологии, являются определяющими факторами в решении фундаментальных проблем, с которыми сталкивается общество.

Таким образом, данное руководство направлено на поддержку планирования соответствующих правил, политики и программ развития человеческого потенциала, чтобы гарантировать, что искусственный интеллект станет инструментом, который действительно приносит пользу и расширяет возможности учителей, учащихся и исследователей. Руководящие принципы, основанные

на Рекомендации ЮНЕСКО об этических аспектах искусственного интеллекта, опираются на человекоориентированный подход, поощряющий человеческую активность, инклюзивность, справедливость, гендерное равенство, культурное и языковое разнообразие, а также плюрализм мнений и их выражение.

В руководстве изначально рассматриваются определение и функционирование искусственного интеллекта, представляя различные доступные технологии и модели (раздел 1), а затем рассматривается ряд дискуссионных этических и политических вопросов, касающихся как искусственного интеллекта в целом, так и его генеративной разновидности (раздел 2). Затем происходит обсуждение шагов и ключевых элементов, которые необходимо учесть при попытке регулирования искусственного интеллекта

в рамках подхода, ориентированного на человека, с целью обеспечения этического, безопасного, справедливого и значимого использования (раздел 3). Далее в разделе 4 предлагаются меры, которые можно предпринять для разработки систематических и всеохватывающих политических рамок для регулирования использования искусственного интеллекта в сфере образования и исследований, а в разделе 5 анализируются возможности творческого использования искусственного интеллекта в разработке учебных программ, преподавании, обучении и исследовательской деятельности. Руководство завершается разделом 6, в котором представлены соображения относительно долгосрочных последствий использования искусственного интеллекта в области образования и исследований.

1. Что такое генеративный ИИ и как он работает?

1.1 Что такое генеративный ИИ?

Генеративный искусственный интеллект (ИИ) представляет собой технологию искусственного интеллекта, которая автоматически создает контент в ответ на запросы, формулируемые в диалоговых интерфейсах на естественном языке. Вместо организации существующих веб-страниц с использованием имеющегося контента генеративный искусственный интеллект фактически осуществляет генерацию нового контента. Он может появляться в форматах, которые включают в себя все символические представления человеческого мышления: тексты, написанные на естественном языке, изображения (включая фотографии, цифровые картины и мультфильмы), видео, музыку и программный код. Генеративный ИИ обучается с использованием данных, полученных из веб-страниц, диалогов в социальных сетях и других источников электронных СМИ. Он осуществляет генерацию контента путем проведения статистического анализа распределения слов, пикселей или других элементов в принятых данных, а также выявления и повторения общих шаблонов, таких как последовательность слов.

Несмотря на то, что генеративный ИИ может создавать новый контент, он не может генерировать новые идеи или решения проблем реального мира, поскольку он не понимает объекты реального мира или социальные отношения, лежащие в основе языка. Более того, несмотря на его быструю и впечатляющую производительность, нельзя полностью доверять генеративному ИИ.

Действительно, даже разработчик ChatGPT признает: «Хотя такие инструменты, как ChatGPT, часто могут генерировать ответы, которые кажутся разумными, нельзя полагаться на их точность». (OpenAI, 2023). Ошибки чаще остаются незамеченными, если у пользователя недостаточные знания темы.

1.2 Как работает генеративный ИИ?

Конкретные технологии, лежащие в основе генеративного ИИ, составляют часть семейства технологий ИИ, известного как машинное обучение (МО). Они используют алгоритмы, которые позволяют постоянно и автоматически улучшать свою производительность на основе данных. Тип МО, который привел ко многим достижениям в области искусственного интеллекта, наблюдаемым в последние годы, таких как использование искусственного интеллекта для распознавания лиц, известен как искусственные нейронные сети (ИНС), основанные на том, как работает человеческий мозг и его синаптические связи между нейронами. Существует много типов искусственных нейронных сетей.

Технологии генерации текста и изображений основаны на наборе технологий искусственного интеллекта, которые были доступны исследователям в течение нескольких лет. Например, ChatGPT использует генеративный предварительно обученный преобразователь (GPT), тогда как генеративный ИИ для изображений обычно использует то, что известно как генеративно-состязательные сети (GAN) (см. **Таблицу 1**)³.

Таблица 1 . Методы, используемые в генеративном искусственном интеллекте

Машинное обучение (МО)		Тип ИИ, который использует данные для автоматического улучшения своей производительности.
Искусственная нейронная сеть (ИНС)		Тип машинного обучения, аналогичный структуре и функционированию человеческого мозга (например, синаптические связи между нейронами).
Текстовый генеративный искусственный интеллект	Трансформеры общего назначения	Тип ИНС, способный концентрироваться на различных частях данных, чтобы определить, как они связаны друг с другом.
	Большие языковые модели (LLM)	Тип преобразователя общего назначения, который обучается на огромных объемах текстовых данных.
	Генеративный предварительно обученный трансформер (GPT)⁴	Тип большой языковой модели, предварительно обученной на обширных данных, повышающих ее способность к восприятию языковых нюансов и созданию качественного текста, адаптированного к конкретному контексту.
Генератор изображений ИИ	Генеративно-состязательные сети (GAN)	Типы нейронных сетей, используемых для генерации изображений.
	Вариационные автокодировщики (VAE)	

1.2.1. Как работают текстовые модели генеративного ИИ

Текстовый генеративный ИИ использует тип ИНС, известный как преобразователь общего назначения, и тип преобразователя общего назначения, называемый большой языковой моделью. Именно поэтому системы ИИ для генерации текста часто называют большими языковыми моделями или LLM. Тип LLM, используемый текстовым генеративным ИИ, известен как генеративный (**G**enerative) предварительно обученный (**P**re-trained) преобразователь (**T**ransformer), или GPT (поэтому используется «GPT» в «ChatGPT»).

ChatGPT построен на GPT-3, разработанном OpenAI. Это была третья итерация GPT: первая была запущена в 2018 году, а самая последняя, GPT-4, в марте 2023 года (см. **Таблицу 2**). Каждый GPT был итеративно улучшен за счет достижений в архитектуре ИИ, методах обучения и методах оптимизации. Одним из хорошо известных аспектов

его непрерывного прогресса является использование растущих объемов данных для обучения экспоненциально растущего числа «параметров». Параметры можно рассматривать как аналоговые регуляторы для точной регулировки работы GPT с целью улучшения производительности. Они включают «весы» модели - числовые параметры, определяющие, как модель обрабатывает ввод и производит вывод данных.

Помимо достижений в оптимизации архитектуры искусственного интеллекта и методов обучения, такая быстрая итерация стала возможной также благодаря огромным объемам данных⁵ и улучшениям вычислительных возможностей, доступных крупным компаниям. С 2012 года вычислительные мощности, используемые для обучения моделей генеративного ИИ, удваиваются каждые 3–4 месяца. Для сравнения, закон Мура имел двухлетний период удвоения (OpenAI, 2018; Стэнфордский университет, 2019).

Таблица 2 . GPT от OpenAI

Модель	Год запуска	Объем обучающих данных	Количество параметров	Характеристики
GPT-1	2018 год	40 ГБ	117 миллионов	Способен выполнять задачи обработки естественного языка, такие как заполнение текстов и ответы на вопросы.
GPT-2	2019 год	40 ГБ	1,500 миллионов	Способен решать более сложные задачи по обработке естественного языка, такие как машинный перевод и обобщение.
GPT-3	2020 год	17 000 ГБ	175 000 миллионов	Способен решать сложные задачи обработки естественного языка, такие как написание связных абзацев и создание целых статей. Также способен адаптироваться к новым задачам всего на нескольких примерах.
GPT-4 ⁶	2023 год	1 000 000 ГБ (сообщается, но не подтверждено)	170 000 000 миллионов (сообщалось, но не подтверждено)	Повышенная надежность и возможность обработки более сложных инструкций.

После того, как GPT был обучен, генерация текстового ответа на запрос включает в себя следующие шаги:

1. Запрос разбивается на более мелкие единицы (называемые токенами), которые вводятся в GPT.
2. GPT использует статистические закономерности для прогнозирования вероятных слов или фраз, которые могут сформировать последовательный ответ на запрос.
 - GPT идентифицирует шаблоны слов и фраз, которые обычно встречаются в его предварительно созданной модели больших данных (которая включает текст, полученный из Интернета и других источников).
 - Используя эти шаблоны, GPT оценивает вероятность появления определенных слов или фраз в данном контексте.
 - Начиная со случайного прогноза, GPT использует эти вероятности, чтобы предсказать следующее вероятное слово или фразу в своем ответе.
3. Предсказанные слова или фразы преобразуются в читаемый текст.
4. Читаемый текст фильтруется через так называемые «защитные ограждения» для удаления любого оскорбительного контента.
5. Шаги 2–4 повторяются до тех пор, пока ответ не будет завершен. Ответ считается завершенным, когда он достигает максимального предела токенов или

соответствует заранее определенным критериям остановки.

6. Ответ дорабатывается для улучшения читаемости путем применения форматирования, знаков препинания и других улучшений (например, начало ответа словами, которые мог бы использовать человек, например «Конечно», «Безусловно» или «Извините»).

Хотя возможность использования GPT и аналогичные возможности для автоматической генерации текста стали доступны исследователям еще с 2018 года, важность запуска ChatGPT заключается в его беспрепятственном доступе через простой в использовании интерфейс. Это означает, что любой, кто имеет доступ к Интернету, может пользоваться этим инструментом. Запуск ChatGPT вызвал шок по всему миру и быстро привел к тому, что другие глобальные технологические и развивающиеся компании вступили в гонку, запуская собственные аналогичные системы, либо создавая новые инструменты на их основе.

К июлю 2023 года были представлены следующие альтернативы ChatGPT:

- **Alpaca:**⁷ доработанная версия большой языковой модели от Meta (Meta's Llama) из Стэнфордского университета, целью которой является борьба с ложной информацией, социальными стереотипами

и токсичным языком выпускников магистратуры.

- **Bard:**⁸ Большая языковая модель от Google, основанная на системах LaMDA и PaLM 2, имеющая доступ к Интернету в режиме реального времени, что означает, что она может предоставлять актуальную информацию.
- **Chatsonic:**⁹ чат, созданный Writesonic, он основан на ChatGPT, а также сканирует данные непосредственно из Гугл (Google). Соответственно, у него меньше шансов дать фактически неправильные ответы.
- **Ernie** (также известный как **Wenxin Yiyao** 文心一言):¹⁰ двуязычная большая языковая модель от Baidu, который все еще находится в разработке и объединяет обширные знания с огромными наборами данных для создания текста и изображений.
- **Hugging Chat:**¹¹ создан компанией Hugging Face, которая подчеркивала этику и прозрачность на протяжении всего процесса разработки, обучения и запуска. Кроме того, все данные, используемые для обучения моделей, имеют открытый исходный код.
- **Jasper:**¹² набор инструментов и API, которые, например, можно обучить писать в предпочитаемом пользователем стиле. Он также может генерировать изображения.
- **Llama:**¹³ Большая языковая модель с открытым исходным кодом от Meta, который требует меньше вычислительной мощности и меньше ресурсов для тестирования новых подходов, проверки работы других и изучения новых вариантов использования.
- **Open Assistant:**¹⁴ подход с открытым исходным кодом, предназначенный для того, чтобы любой человек, обладающий достаточным опытом, мог разработать свою собственную большую языковую модель. Он был построен на данных обучения, собранных волонтерами.
- **Tongyi Qianwen** (通义千问):¹⁵ Большая языковая модель от Alibaba, которая

может отвечать на запросы на английском или китайском языках. Она интегрируется в набор бизнес-инструментов Alibaba.

- **YouChat:**¹⁶ Большая языковая модель, которая включает возможности поиска в реальном времени для предоставления дополнительного контекста и информации для получения более точных и надежных результатов.

Большинство из них можно использовать бесплатно, хотя с ограничениями, и некоторые обладают открытым исходным кодом. Также существует множество других продуктов, основанных на одной из вышеупомянутых больших языковых моделей, например:

- **ChatPDF:**¹⁷ обобщает и отвечает на вопросы по отправленным PDF-документам.
- **Elicit: Помощник по исследованиям искусственного интеллекта:**¹⁸ предназначен для автоматизации частей рабочих процессов исследователей, выявления соответствующих статей и обобщения ключевой информации.
- **Perplexity:**¹⁹ обеспечивает «базу знаний» для людей, ищущих быстрые и точные ответы, соответствующие их потребностям.

Аналогичным образом, инструменты на основе больших языковых моделей встраиваются в другие продукты, например, в веб-браузеры. Ниже представлены расширения для браузера Chrome, созданные на основе ChatGPT:

- **WebChatGPT:**²⁰ предоставляет ChatGPT доступ в Интернет для более точного и актуального общения.
- **Compose AI:**²¹ автозаполняет предложения в электронных письмах и других местах.
- **TeamSmart AI:**²² предоставляет «команду виртуальных помощников».
- **Wiseone:**²³ упрощает получение информации в Интернете.

Кроме того, ChatGPT был включен в некоторые поисковые системы²⁴ и внедряется в большие портфели инструментов

повышения производительности (например, Microsoft Word и Excel), что делает его еще более доступным в офисах и образовательных учреждениях по всему миру (Мерфи и Келли, 2023).

Наконец, в качестве интересного перехода к созданию изображений с помощью генеративного ИИ, стоит отметить, что последняя версия GPT от OpenAI, а именно GPT-4, обладает способностью обрабатывать запросы не только в виде текста, но и в формате изображений. В этом смысле он мультимодален. Соответственно, некоторые утверждают, что название «большая языковая модель» (LLM) становится менее подходящим, и это одна из причин, по которой исследователи из Стэнфордского университета предложили термин «базовая модель» (Боммасани и др., 2021). Эта альтернатива еще не получила широкого распространения.

1.2.2. Как работают модели изображений генеративного ИИ

Изображение и музыка, созданные генеративным ИИ, обычно используют другой тип ИНС, известный как генеративно-состязательные сети (GAN), который также можно комбинировать с вариационными автокодировщиками. GAN состоят из двух частей (два «противника»): «генератор» и «дискриминатор». В случае GAN изображений генератор создает случайное изображение в ответ на запрос, а дискриминатор пытается отличить это сгенерированное изображение от реальных изображений. Затем генератор использует результат дискриминатора для настройки своих параметров и создания другого изображения. Процесс повторяется, возможно, тысячи раз, при этом генератор создает все более и более реалистичные изображения, которые дискриминатор все менее и менее способен отличить от реальных изображений. Например, успешная GAN, обученная на наборе данных из тысяч фотографий пейзажей, может генерировать новые, но нереальные изображения пейзажей, которые почти неотличимы от реальных фотографий. Между тем, GAN, обученная на наборе данных популярной музыки (или даже музыки одного

исполнителя), может генерировать новые музыкальные произведения, повторяющие структуру и сложность оригинальной музыки.

По состоянию на июль 2023 года доступны следующие модели **генеративного искусственного интеллекта для создания изображений, все из которых генерируют изображения из текстовых подсказок**. Большинство из них можно использовать бесплатно, с ограничениями.

- **Craiyon**:²⁵ ранее известный как DALL·E mini.
- **DALL·E 2**:²⁶ инструмент генеративного ИИ от OpenAI.
- **DreamStudio**:²⁷ инструмент генеративного ИИ для изображений Stable Diffusion.
- **Fotor**:²⁸ включает генеративный ИИ в ряд инструментов для редактирования изображений.
- **Midjourney**:²⁹ независимый инструмент генеративного ИИ для изображений.
- **NightCafe**:³⁰ интерфейс для Stable Diffusion и DALL·E 2.
- **Photosonic**:³¹ генератор искусственного интеллекта WriteSonic.

Примеры доступного **искусственного интеллекта для создания видео** включают в себя следующее:

- **Elai**:³² Может конвертировать презентации, веб-сайты и текст в видео.
- **GliaCloud**:³³ может генерировать видео на основе новостного контента, публикаций в социальных сетях, прямых трансляций спортивных событий и статистических данных.
- **Pictory**:³⁴ может автоматически создавать короткие видеоролики из длинного контента.
- **Runway**:³⁵ предлагает ряд инструментов для создания и редактирования видео (и изображений).

Наконец, вот несколько примеров доступного **генеративного ИИ для создания музыки**:

- **Aiva**:³⁶ может автоматически создавать персонализированные саундтреки.
- **Boomy**,³⁷ **Soundraw**,³⁸ и **Voicemod**:³⁹ могут создавать песни из любого текста и не требуют знаний музыкальной композиции.

1.3 Оптимизация запросов для достижения желаемых результатов

Несмотря на то, что использование генеративного искусственного интеллекта может показаться очень простым - ввод вопроса или запроса, на практике пользователю все еще нелегко получить именно тот результат, который он хочет. Например, революционное изображение с помощью искусственного интеллекта *Théâtre D'opéra Spatial*, получившее приз на выставке штата Колорадо в Соединенных Штатах Америки (Colorado State Fair), потребовало нескольких недель написания запросов и точной настройки сотен изображений, чтобы сгенерировать окончательную версию (Руз, 2022). Аналогичная задача по написанию эффективных запросов для текстового генеративного ИИ привела к тому, что на веб-сайтах по подбору персонала появляется все больше вакансий по разработке запросов (Попли, 2023). «Оптимизация запросов» означает процессы и методы формирования входных данных для генеративного искусственного интеллекта, обеспечивая более точные результаты, соответствующие намерениям пользователя.

Разработка успешных запросов осуществляется путем формулирования последовательной цепочки объяснений, сосредоточенной на конкретной задаче, или логической последовательности мыслей. Конкретные рекомендации включают в себя:

- Используйте **простой** и понятный язык, избегая сложных или двусмысленных формулировок.

- Включите **примеры**, иллюстрирующие желаемый результат или формат создаваемых дополнений.
- Включите **контекст**, который имеет решающее значение для создания релевантных и содержательных дополнений.
- **Уточняйте** и повторяйте по мере необходимости, экспериментируя с различными вариантами.
- **Соблюдайте принципы** этики, избегая запросов, которые могут привести к созданию неприемлемого, предвзятого или **наносщего вред контента**.

Также важно отметить, что на результаты, выданные генеративным ИИ, нельзя полагаться без критической оценки. Например, OpenAI пишет о своем самом сложном GPT:⁴⁰

«Несмотря на возможности, у GPT-4 все еще присутствуют те же ограничения, что и у предыдущих моделей GPT. Она по-прежнему не является полностью надежной моделью (иногда создает ложные факты и допускает ошибки в логике). При использовании выходных данных этой языковой модели следует проявлять осторожность, особенно если она используется в ситуациях, где важны последствия. При этом точный протокол (например, проверка человеком, обоснование с дополнительным контекстом или полное избегание использования GPT в случаях, когда очень важны последствия) должен соответствовать потребностям конкретного сценария использования.»

Прежде чем применять инструменты генеративного ИИ в масштабных и критически важных сценариях, необходимо провести строгие пользовательские тесты и оценки производительности, учитывая качество результатов, которые они предоставляют. Эти оценочные упражнения должны основываться на показателях производительности, которые напрямую соответствуют типу задач, для которых пользователи намерены использовать генеративный ИИ. Например, при решении математических задач ключевым показателем может стать «точность», отражающая, насколько часто инструмент генеративного ИИ дает верные ответы; в случае ответов на деликатные вопросы, решающим может быть

«скорость ответов» (как быстро генеративный ИИ предоставляет ответы на вопросы); для генерации кода важным критерием может стать «доля сгенерированных кодов, которые могут быть непосредственно исполнимы» (может ли сгенерированный код быть применен в программировании и успешно пройти модульное тестирование); для изображений определяющим может быть «точное соответствие» (насколько точно сгенерированные визуальные объекты соответствуют действительности) (Чен, 2023).

В заключение, можно сказать, что на первый взгляд генеративный ИИ может показаться простым в использовании; однако для более сложных задач требуется профессиональное участие человека и критическая оценка перед использованием результатов.



Последствия для образования и научных исследований:

Несмотря на то, что генеративный ИИ может помочь учителям и исследователям создавать полезные тексты и другие материалы для поддержки своей работы, этот процесс не всегда прост. Для достижения желаемого результата может понадобиться несколько запросов с последующими уточнениями. Особенно беспокойство вызывает то, что учащиеся, по определению менее опытные, чем их учителя, могут неосознанно без критической оценки принимать результаты генеративного ИИ, которые могут быть поверхностными, неточными или даже вредными.

1.4 Новый EdGPT и его последствия

Учитывая, что модели генеративного ИИ могут служить основой или отправной точкой для разработки более специализированных или предметно-ориентированных моделей, некоторые исследователи предложили переименовать GPT в «базовые модели» (Боммасани, 2021). В сфере образования разработчики и исследователи начали

дорабатывать базовую модель для создания EdGPT (от англ. Education – образование + GPT).⁴¹ Модели EdGPT обучаются на конкретных данных для образовательных целей. Другими словами, EdGPT стремится усовершенствовать модель, полученную на основе огромных объемов данных общего обучения, с меньшими объемами высококачественных данных об образовании, специфичных для конкретной предметной области.

Это потенциально дает EdGPT больше возможностей для поддержки преобразований, перечисленных в разделе 4.3. Например, модели EdGPT, нацеленные на совместную разработку учебных программ, могут позволить преподавателям и учащимся создавать соответствующие учебные материалы, такие как планы уроков, викторины и интерактивные занятия, которые тесно связаны с эффективным педагогическим подходом и конкретными учебными целями и уровнями сложности для определенных учащихся. Аналогичным образом, в контексте индивидуального тренера по языковым навыкам базовая модель, дополненная текстами, подходящими для конкретного языка, может использоваться для создания образцовых предложений, абзацев или диалогов для практики. Когда учащиеся взаимодействуют с моделью, она может ответить релевантным и грамматически точным текстом на нужном для них уровне. Теоретически выходные данные моделей EdGPT также могут содержать меньше общих ошибок или иного нежелательного контента, чем стандартный GPT, но все же могут выдавать ошибки. Крайне важно отметить, что, если базовые модели и подход к генеративному ИИ существенно не изменятся, EdGPT все равно может выдавать ошибки и демонстрировать другие ограничения. Соответственно, по-прежнему важно, чтобы основные пользователи EdGPT, особенно преподаватели и учащиеся, критически относились к любым результатам.

В настоящее время доработка базовых моделей для более целевого использования GPT в образовании находится на ранней стадии. Существующие примеры включают EduChat, базовую модель, разработанную Восточно-Китайским педагогическим университетом

для предоставления услуг по преподаванию и обучению, коды, данные и параметры которого доступны в виде открытого исходного кода.⁴² Другим примером является MathGPT, разрабатываемый TAL Education Group, большая языковая модель, которая занимается решением математических задач и чтением лекций для пользователей по всему миру.⁴³

Однако, прежде чем станет возможен значительный прогресс, важно, чтобы усилия были направлены на совершенствование базовых моделей не только путем добавления предметных знаний и устранения предвзятости, но также путем добавления знаний о соответствующих методах обучения и о том, как это может быть отражено в разработке алгоритмов и модели.

Задача состоит в том, чтобы определить, в какой степени модели EdGPT могут выходить за рамки предметных знаний и ориентироваться на педагогику, ориентированную на учащихся, и на позитивное взаимодействие учителя и учащихся. Дальнейшая задача состоит в том, чтобы определить степень, в которой данные об учащихся и учителях могут быть этически собраны и использованы для улучшения EdGPT. Наконец, существует также необходимость в серьезных исследованиях, чтобы гарантировать, что EdGPT не нарушает права учащихся и полномочия учителей.

2. Разногласия по поводу генеративного ИИ и их значение для образования

В предыдущем разделе обсуждалось, что такое генеративный искусственный интеллект (ИИ) и как он функционирует. В данной части рассматриваются противоречия и этические риски, связанные со всеми системами генеративного ИИ, а также освещаются некоторые последствия для образования.

2.1 Усугубление цифровой бедности

Как отмечалось ранее, генеративный искусственный интеллект (ИИ) требует огромных объемов данных, значительной вычислительной мощности, а также непрерывных инноваций в структуре искусственного интеллекта и методах обучения. Подобные ресурсы в основном доступны крупнейшим международным технологическим компаниям и некоторым экономикам, в основном США, Китаю и, в меньшей степени, Европе. Это означает, что возможность создания и управления генеративным ИИ в значительной степени недоступна для большинства компаний и стран, особенно для стран, относящихся к Глобальному Югу.

Поскольку доступ к данным становится все более важным для экономического развития стран и цифровых возможностей отдельных лиц, те страны и граждане, которые лишены доступа к данным или не могут позволить себе их использование, остаются в ситуации «информационной бедности» (Марвала, 2023). Аналогичная ситуация складывается с доступом к вычислительным мощностям. Быстрое распространение генеративного ИИ в технологически развитых странах и регионах экспоненциально ускорило создание и обработку данных и одновременно усилило концентрацию богатства ИИ на Глобальном Севере. Это привело к дополнительному исключению регионов с ограниченным доступом к данным и увеличило их долгосрочный риск подчинения стандартам, заложенным в модели GPT.

Текущие модели ChatGPT обучаются на данных онлайн-пользователей, отражающих ценности и нормы Глобального Севера. Это делает их неприменимыми для локально ориентированных алгоритмов искусственного интеллекта в регионах с ограниченным доступом к данным, преимущественно во многих странах Глобального Юга, а также в менее обеспеченных сообществах Глобального Севера.



Последствия для образования и научных исследований:

Исследователи, преподаватели и обучающиеся должны осуществлять критический анализ ценностных предпосылок, культурных стандартов и социальных обычаев, внедренных в образовательные модели генеративного искусственного интеллекта. Политики должны осознавать и предпринимать меры по решению проблемы усиления неравенства, вызванного увеличением разрыва в обучении и контроле моделей генеративного искусственного интеллекта.

2.2 Опережение адаптации национального регулирования

Доминирующих поставщиков генеративного ИИ также критиковали за то, что они не позволяли своим системам подвергаться строгой независимой академической проверке (Дживеди, 2023).⁴⁴ Основополагающие технологии генеративного ИИ, как правило, охраняются как корпоративная интеллектуальная собственность. Между тем, многим компаниям, которые начинают использовать генеративный ИИ, становится все труднее поддерживать безопасность своих систем (Лин, 2023). Более того, несмотря на призывы к регулированию со стороны самой

отрасли ИИ,⁴⁵ разработка законодательства о создании и использовании всех видов ИИ, включая генеративный ИИ, зачастую отстает от быстрых темпов развития. Это частично объясняет проблемы, с которыми сталкиваются национальные и местные органы при понимании и регулировании правовых и этических вопросов.⁴⁶

Несмотря на то, что генеративный ИИ может расширить человеческие возможности при выполнении определенных задач, демократический контроль над компаниями, продвигающими генеративный ИИ, ограничен. Это поднимает вопрос о регулировании, в частности, в отношении доступа и использования внутренних данных, включая данные о местных учреждениях и отдельных лицах, а также данные, полученные на территории стран. Необходимо соответствующее законодательство, чтобы местные правительственные учреждения могли получить некоторый контроль над нарастающим влиянием генеративного ИИ и обеспечить его управление как общественное благо.



Последствия для образования и научных исследований:

Исследователи, преподаватели и учащиеся должны осознавать отсутствие соответствующих норм и правил для защиты собственности национальных учреждений и частных лиц, а также прав внутренних пользователей генеративного ИИ. Кроме того, они должны быть готовы к реагированию на законодательные проблемы, вызванные внедрением генеративного ИИ.

2.3 Использование контента без согласия владельцев данных

Как уже отмечалось ранее, модели генеративного ИИ строятся на основе обширных объемов данных (таких как текст, звуки, код и изображения), которые часто извлекаются из интернета и, как правило, без разрешения владельца.

В результате многие визуальные и кодовые системы генеративного искусственного интеллекта подверглись обвинениям в нарушении прав интеллектуальной собственности. На момент написания данной публикации ведется несколько международных судебных процессов, касающихся этого вопроса.

Кроме того, некоторые отметили, что GPT могут противоречить таким законам, как Общий регламент по защите данных Европейского Союза (GDPR, 2016 год), особенно в контексте права людей на удаление информации о себе, поскольку в настоящее время невозможно удалить чьи-либо данные (или результаты обработки этих данных) из модели GPT после их внесения для обучения GPT.



Последствия для образования и научных исследований:

Исследователи, преподаватели и учащиеся должны быть осведомлены о правах владельцев данных и осуществлять проверку, чтобы удостовериться, что используемые ими инструменты генеративного ИИ соблюдают существующие нормы и правила.

Кроме того, они должны знать, что изображения или коды, созданные с использованием генеративного ИИ, могут нарушать чьи-либо права интеллектуальной собственности, и осознавать, что изображения, звуки или код, которые они создают и выкладывают в Интернет, могут быть использованы другими системами генеративного ИИ.

2.4 Особенности «необъяснимых» моделей в процессе генерации результатов

Давно признано, что искусственные нейронные сети (ИНС) обычно представляют собой «черные ящики»; то есть их внутренняя работа закрыта для проверки. В результате ИНС не являются «прозрачными» или «объяснимыми», и невозможно установить, как были определены их выходные данные.

Хотя общий подход, включая используемые алгоритмы, в целом объясним, конкретные модели и их параметры, включая вес моделей, не подлежат проверке, поэтому конкретный генерируемый результат не может быть объяснен. В такой модели, как GPT-4 (см. **Таблицу 2**), существуют миллиарды параметров/веса, и именно вес в совокупности содержит изученные шаблоны, которые модель использует для генерации своих результатов. Ввиду непрозрачности параметров/весов в искусственных нейронных сетях (см. **Таблицу 1**) становится невозможным предоставить точные пояснения относительно конкретного метода генерации результатов с использованием этих моделей.

Отсутствие прозрачности и объяснимости в генеративном ИИ становятся все более проблематичным по мере того, как генеративный ИИ становится все более сложным (см. **Таблицу 2**), часто производя неожиданные или нежелательные результаты. Кроме того, модели генеративного ИИ передают и усиливают присутствующие в обучающих данных предубеждения, которые, учитывая не прозрачную природу моделей, сложно обнаружить и устранить. Наконец, эта непрозрачность также является основной причиной проблем доверия к генеративному ИИ (Назарецкий и др., 2022a).



Значение для образования и научных исследований:

Исследователи, преподаватели и учащиеся должны понимать, что системы генеративного ИИ работают как черные ящики и, следовательно, трудно, если вообще возможно, понять, почему был выдан тот или иной результат. Отсутствие объяснения того, как генерируются выходные данные, приводит к тому, что пользователи остаются в рамках логики, определяемой параметрами, разработанными в системах генеративного ИИ. Эти параметры могут отражать конкретные культурные или коммерческие ценности и нормы, которые косвенно искажают создаваемый контент.

Если пользователи не понимают, как модель генеративного ИИ пришла к конкретному результату, они менее склонны принимать ее или использовать (Назарецкий и др., 2022б).

2.5 Искусственный интеллект и загрязнение контента в сети

Поскольку данные обучения GPT обычно берутся из Интернета, который слишком часто содержит дискриминационные и другие неприемлемые выражения, разработчикам приходится внедрять так называемые «ограждения», чтобы выходные данные GPT не были оскорбительными и/или неэтичными. Однако из-за отсутствия строгих правил и эффективных механизмов мониторинга предвзятые материалы, генерируемые искусственным интеллектом, все чаще распространяются по Интернету, нанося вред одному из ключевых источников контента или знаний для большинства обучающихся во всем мире. Это особенно важно, поскольку материал, генерируемый ИИ, может оказаться весьма точным и убедительным, хотя зачастую он содержит ошибки и предвзятые идеи. Это представляет высокий риск для учащихся, которые не имеют глубоких предварительных знаний по рассматриваемой теме. Это также влечет за собой рекурсивный риск для последующих моделей GPT, которые будут обучаться на тексте, извлеченном из Интернета и сформированном самими моделями GPT, содержащим предвзятости и ошибки.



Последствия для образования и исследований:

- Исследователи, преподаватели и учащиеся должны осознавать, что системы генеративного ИИ могут публиковать оскорбительные и неэтичные материалы.
- Необходимо также осознавать долгосрочные проблемы, которые могут возникнуть в контексте надежности знаний, когда будущие модели GPT будут базироваться на тексте, сгенерированном предыдущими моделями GPT.

2.6 Недостаточное понимание реального мира

Текстовые GPT иногда пренебрежительно называют «стохастическими попугаями». Это связано с тем, что несмотря на то, что они способны создавать текст, который может казаться убедительным, в них часто содержатся ошибки и вредные утверждения (Бендер и др., 2021). Эту тенденцию обуславливает факт, что модели GPT лишь репродуцируют языковые шаблоны, выявленные в своих обучающих данных (как правило, это текст, собранный из Интернета), начиная с случайных (или «стохастических») шаблонов, и не обладают пониманием их значения — аналогично тому, как попугай может имитировать звуки, но при этом не осознавая их смыслового содержания.

Расхождение между моделями генеративного искусственного интеллекта, которые «демонстрируют» понимание используемого ими и создаваемого текста, и реальностью, заключающейся в их отсутствии понимания языка и реального мира, может привести к ситуации, когда преподаватели и учащиеся доверяют результатам, производимым этими моделями, но не могут гарантировать их качество. Это создает серьезные риски для будущего образования. Фактически, генеративный ИИ не базируется на наблюдениях реального мира или других ключевых аспектах научного метода, и не соответствует человеческим или социальным ценностям. По этим причинам он не может создавать по-настоящему новый контент о реальном мире, объектах и их взаимоотношениях, людях и социальных связях, отношениях между человеком и объектом или между человеком и технологиями. Вопрос о том, можно ли считать явно новый контент, созданный моделями генеративного ИИ, научным знанием, остается спорным.

Как уже отмечалось, GPT часто может создавать неточный или неправдоподобный текст. На самом деле хорошо известно, что GPT создает некоторые вещи, которые не существуют в реальной жизни. Некоторые называют это «галлюцинацией», хотя другие критикуют использование такого антропоморфного и, следовательно, вводящего в заблуждение термина. Это признают компании-производители генеративного ИИ. Например, в нижней

части страницы ChatGPT указано: «ChatGPT может предоставлять неточную информацию о людях, местах или фактах».²

Некоторые сторонники также высказывают предположение о том, что генеративный искусственный интеллект представляет собой значительный шаг в направлении общего искусственного интеллекта (AGI — термин, описывающий класс искусственного интеллекта, более развитого в умственных способностях, чем человек). Тем не менее, данное утверждение давно подвергается критике с предположением, что искусственный интеллект, по крайней мере на данный момент, не сможет достичь уровня общего искусственного интеллекта (AGI), если он не сможет каким-то образом объединить в симбиозе искусственный интеллект, основанный на знаниях (также известный как символический или основанный на правилах ИИ), и искусственный интеллект, основанный на данных (также известный как машинное обучение). (Маркус, 2022). Утверждения о возможности достижения общего искусственного интеллекта (AGI) или разума также отвлекают наше внимание от более тщательного рассмотрения текущего вреда, вызываемого искусственным интеллектом, такого как скрытая дискриминация групп, которые уже подверглись дискриминации. (Мец, 2021).



Последствия для образования и исследований:

- Текст, созданный генеративным ИИ, может визуально напоминать человеческий. Однако следует отметить, что генеративный ИИ не обладает пониманием. Он, скорее, объединяет слова, используя шаблоны, распространенные в сети. Таким образом, создаваемый текст может содержать ошибки.
- Исследователи, преподаватели и учащиеся должны осознавать, что GPT не понимает текст, который он генерирует; что он может создавать зачастую неверные утверждения; и, следовательно, им необходимо критически подходить ко всему, что создано генеративным ИИ.

2.7 Ограничение разнообразия мнений и дальнейшая маргинализация голосов

Инструменты, такие как ChatGPT, обычно проявляют тенденцию к генерации ответов, которые отражают предполагаемые смысловые значения, заложенные владельцами данных или создателями, примененными в процессе обучения этих моделей. В сущности, если последовательность слов часто встречается в тренировочных данных – как в случае общепринятых и непротиворечивых тем или основополагающих убеждений – она, вероятнее всего, будет воспроизведена GPT в выходных данных.

Это создает потенциальную угрозу для ограничения и подрыва развития плюрализма мнений и разнообразного выражения идей. Группы населения, имеющие ограниченный доступ к данным, включая маргинализированные сообщества в мировом масштабе, часто имеют минимальное или ограниченное цифровое

присутствие в сети. Вследствие недостаточного учета и внимания к голосам определенных групп в данных, применяемых для обучения моделей, таких как GPT, их позиции недостаточно отражены, что в итоге приводит к ограниченному представлению их точек зрения в сгенерированных результатах. Учитывая методологию предварительного обучения, основанную на данных с веб-страниц и социальных сетей, модели GPT могут углублять маргинализацию уже уязвимых групп населения.

2.8 Создание более реалистичных дипфейков

Помимо основных противоречий, присутствующих всему генеративному искусственному интеллекту, технология GAN (генеративно-состязательные сети) в генеративном искусственном интеллекте может использоваться для модификации или манипуляции существующими изображениями или видео с целью создания фальшивых материалов, которые сложно отличить от реальных. Это способствует созданию «дипфейков» и так называемых «фейковых новостей». Другими словами, генеративный искусственный интеллект облегчает определенным субъектам возможность осуществлять неэтичные, аморальные и преступные действия, включая распространение дезинформации, проведение пропаганды разжигания ненависти и создание полностью ложных, а порой и компрометирующих материалов с участием людей без их согласия и ведома.



Последствия для образования и исследований:

- Несмотря на то, что разработчики и поставщики моделей генеративного ИИ несут основную ответственность за постоянное устранение систематических ошибок в наборах данных и выходных данных этих моделей, исследователи, преподаватели и учащиеся на стороне пользователей должны знать, что выходные данные представляют собой только наиболее распространенный или доминирующий взгляд на мир в то время, когда были получены данные обучения, и что некоторые из них являются ошибочными или предвзятыми (например, стереотипные гендерные роли).
- Учащиеся, преподаватели и исследователи должны всегда подходить к информации, предоставляемой искусственным интеллектом, с критическим взглядом.
- Исследователи, преподаватели и учащиеся также должны осознавать, что голоса меньшинств могут быть не представлены, поскольку голоса меньшинств по определению менее распространены в данных обучения.



Последствия для образования и исследований:

Хотя поставщики генеративного ИИ обязаны защищать авторские права пользователей, исследователи, преподаватели и учащиеся также должны осознавать, что любые изображения, которыми они делятся в Интернете, могут быть включены в обучающие данные генеративного ИИ, подвергнуты манипуляциям и использованы в неэтичных целях.

3. Регулирование использования генеративного ИИ в образовании

Чтобы разрешить противоречия вокруг генеративного ИИ и использовать потенциальные преимущества генеративного ИИ в образовании, его необходимо регулировать. Регулирование генеративного ИИ в образовательных целях требует ряда шагов и политических мер, основанных на человеко-ориентированном подходе, для обеспечения его этического, безопасного, справедливого и значимого использования.

3.1 Ориентированный на человека подход к искусственному интеллекту

«Рекомендация ЮНЕСКО об этических аспектах искусственного интеллекта» (2021 год) предоставляет необходимую нормативную основу для разрешения многочисленных противоречий вокруг генеративного ИИ, включая те, которые связаны с образованием и исследованиями. Она основана на гуманоцентрическом подходе к искусственному интеллекту, ставящем в центр развитие человеческого потенциала для достижения инклюзивного, справедливого и устойчивого будущего. Данный подход опирается на принципы прав человека и необходимость защиты человеческого достоинства и культурного разнообразия, определяющих общественные ценности. С управленческой точки зрения человеко-ориентированный подход требует соответствующего регулирования, которое способствует человеческой активности, прозрачности и отчетности перед обществом.

В «Пекинском Консенсусе по искусственному интеллекту (ИИ) и образованию» (2019 год) дополнительно уточняется, что подразумевает человеко-ориентированный подход к использованию ИИ в контексте образования. Консенсус подтверждает, что использование технологий ИИ в образовании должно повысить человеческий потенциал для

устойчивого развития и эффективного сотрудничества человека и машины в жизни, обучении и работе. Он также призывает к дальнейшим действиям по обеспечению равного доступа к ИИ для поддержки маргинализированных групп населения и устранения неравенства, одновременно содействуя языковому и культурному разнообразию. Консенсус предлагает принять общеправительственные, межсекторальные и многосторонние подходы к планированию политики использования ИИ в образовании.

«Технологии искусственного интеллекта в образовании: перспективы и последствия: Руководство для лиц, ответственных за формирование политики» (ЮНЕСКО, 2022б) дополнительно уточняет, что означает человеко-ориентированный подход при изучении преимуществ и рисков ИИ в образовании, а также роли образования как средства развития компетенций ИИ. В документе содержатся конкретные рекомендации по формулированию политики в отношении использования искусственного интеллекта с целью:

- (i) обеспечения инклюзивного доступа к образовательным программам, особенно для уязвимых групп, таких как учащиеся с ограниченными возможностями;
- (ii) поддержки персонализированных и открытых форм обучения;
- (iii) улучшения управления на основе данных для расширения доступа и повышения качества обучения;
- (iv) контроля образовательных процессов и предупреждения учителей о рисках неудач;
- (v) развития понимания и навыков этичного и целенаправленного использования искусственного интеллекта.

3.2 Шаги по регулированию генеративного ИИ в образовании

До выпуска ChatGPT правительства разрабатывали или адаптировали основы

регулирования сбора и использования данных, а также внедрения систем ИИ в различных секторах, в том числе в образовании, что обеспечивало законодательный и политический контекст для регулирования новых приложений ИИ. После выпуска нескольких конкурирующих моделей генеративного ИИ, начиная с ноября 2022 года, правительства принимают различные политические меры – от запрета генеративного ИИ до оценки потребностей в адаптации существующих рамок и срочной разработки новых правил.

Государственные стратегии регулирования и содействия творческому использованию генеративного ИИ были рассмотрены в апреле 2023 года (ЮНЕСКО, 2023б).⁴⁷ В обзоре предлагается серия из семи шагов, которые правительственные учреждения могут предпринять для регулирования генеративного ИИ и восстановления общественного контроля, чтобы использовать его потенциал в различных секторах, в том числе в образовании.

Шаг 1. Утвердить международные или региональные общие правила защиты данных или разработать национальные правила.

Обучение моделей генеративного ИИ включает сбор и обработку онлайн-данных от граждан многих стран. Использование данных и контента без согласия еще больше усложняет вопрос защиты личных данных.

Например, Общий регламент по защите данных Европейского Союза, принятый в 2018 году, обеспечивает необходимую правовую основу для регулирования сбора и обработки персональных данных поставщиками генеративного ИИ. В соответствии с данными портала «Законодательство о защите данных и конфиденциальности в мире» и информацией, представленной на Конференции Организации Объединенных Наций по торговле и развитию (ЮНКТАД), из 194 стран 137 внедрили законодательные меры, направленные на обеспечение защиты данных и конфиденциальности.⁴⁸

Однако степень реализации этих рамок в этих странах остается неясной. Поэтому как никогда

важно обеспечить их правильную реализацию, включая регулярный мониторинг работы систем генеративного ИИ. Странам, в которых еще нет общих законов о защите данных, также необходимо срочно их разработать.

Шаг 2. Принять/пересмотреть финансирование общегосударственных стратегий в области ИИ.

Регулирование генеративного ИИ должно быть неотъемлемой частью более широких национальных стратегий ИИ, которые могут обеспечить безопасное и справедливое использование ИИ во всех секторах развития, в том числе в образовании. Формулирование, одобрение, финансирование и реализация национальных стратегий в области ИИ требуют общегосударственного подхода. Только такой подход может обеспечить координацию межсекторальных действий, необходимых для комплексного реагирования на возникающие вызовы.

К началу 2023 года около 67 стран⁴⁹ разработали или запланировали национальные стратегии по ИИ, причем 61 из них приняли форму отдельной стратегии ИИ, а 7 представляли собой главы по ИИ, интегрированные в более широкие национальные стратегии в области ИКТ или цифровизации. Исходя из того, что генеративный ИИ является достаточно новым явлением, следует отметить, что до настоящего времени ни одна из рассматриваемых национальных стратегий не включала генеративный ИИ как конкретный объект внимания.

Крайне важно, чтобы страны пересмотрели существующие национальные стратегии в области ИИ или разработали их, обеспечив положения, регулирующие этическое использование ИИ во всех секторах, включая образование.

Шаг 3. Утвердить и внедрить конкретные правила этики ИИ.

Для решения этических вопросов, связанных с использованием ИИ, необходимы специальные правила.

Обзор существующих национальных стратегий ИИ, проведенный ЮНЕСКО в 2023 году, показывает, что выявление таких этических проблем и формулирование руководящих принципов является общим только для примерно 40 национальных стратегий ИИ.⁵⁰ Требуется внедрить этические принципы в действующие законы или правила, и это происходит редко. Действительно, только около 20 стран определили какие-либо четкие правила этики ИИ, в том числе в отношении образования, либо в рамках национальных стратегий ИИ, либо иным образом. Интересно, что, хотя образование выделено в качестве области политики примерно в 45 национальных стратегиях ИИ,⁵¹ упоминания об образовании больше сформулированы с точки зрения навыков ИИ и развития талантов, необходимых для поддержки национальной конкурентоспособности, и меньше с точки зрения этических вопросов.

Страны, в которых еще нет правил этики ИИ, должны сформулировать и внедрить их как можно скорее.

Шаг 4. Скорректировать или закрепить существующие законы об авторском праве, чтобы регулировать контент, созданный искусственным интеллектом.

Все более широкое использование генеративного ИИ создало новые проблемы для авторского права, касающиеся как защищенного авторским правом контента или работы, на которых обучаются модели, так и статуса «нечеловеческих» результатов знаний, которые они производят.

В настоящее время только Китай, страны ЕС и США скорректировали законы об авторском праве с учетом последствий генеративного ИИ. Бюро авторских прав США, например, постановило, что результаты систем генеративного ИИ, таких как ChatGPT, не подлежат защите в соответствии с законодательством США об авторском праве, утверждая, что «авторское право может защитить только материал, который является продуктом человеческого творчества» (Бюро авторских прав США, 2023). Тем временем в ЕС предлагаемый Закон ЕС об искусственном интеллекте требует от разработчиков

инструментов ИИ раскрывать защищенные авторским правом материалы, которые они использовали при создании своих систем (Европейская комиссия, 2021). Китай в своем постановлении о генеративном ИИ, опубликованном в июле 2023 года, требует маркировки результатов контента, созданного генеративным ИИ, и признает их только как результаты цифрового синтеза.

Регулирование использования авторских материалов в процессе обучения генеративных искусственных интеллектуальных моделей, а также определение статуса авторских прав на результаты, сгенерированные генеративным ИИ, представляют собой новые аспекты, требующие внимания в рамках сферы законодательства об авторском праве. Этим обусловлена неотложная необходимость внести коррективы в существующую нормативную базу с учетом этих изменений.

Шаг 5. Разработать нормативно-правовую базу для генеративного ИИ

С развитием технологий ИИ современные национальные и местные органы власти вынуждены ускорить процесс обновления своих правил. По состоянию на июль 2023 года только одна страна – Китай, опубликовала конкретные официальные предписания относительно использования генеративного ИИ. Изданное 13 июля 2023 года (Управление киберпространством Китая, 2023) временное положение по работе службы генеративного искусственного интеллекта Управление киберпространством Китая требует, чтобы поставщики систем генеративного ИИ должным образом и законно маркировали созданный ими контент, изображения и видео в соответствии с действующим положением о глубоком синтезе в рамках онлайн-информационных услуг. Необходимо разработать больше подобных национальных рамок, ориентированных на генеративный ИИ, основываясь на выявленных пробелах в оценке существующих местных правил и законов.

Шаг 6. Нарращивать потенциал для правильного использования генеративного ИИ в образовании и исследованиях.

Школы и другие образовательные учреждения должны развивать потенциал, чтобы понимать потенциальные преимущества и риски ИИ, включая генеративный ИИ, для образования. Только на основе такого понимания они могут оправдать внедрение инструментов ИИ. Более того, учителям и исследователям необходимо оказывать поддержку в укреплении их потенциала по правильному использованию генеративного ИИ, в том числе посредством обучения и непрерывного обучения. Ряд стран запустили такие программы по наращиванию потенциала, в том числе Сингапур, который предлагает специальную платформу для развития потенциала ИИ в образовательных учреждениях через Правительственный облачный кластер ИИ, который включает в себя специальный репозиторий моделей GPT (Окампо, 2023).

Шаг 7. Предусмотреть долгосрочные последствия генеративного ИИ для образования и исследований.

Влияние существующих версий генеративного ИИ только начинает проявляться, требуя более глубокого исследования и осмысления их воздействия на образовательные процессы. Между тем продолжают разрабатываться и внедряться более мощные версии генеративного ИИ и других методов искусственного интеллекта. Тем не менее, остаются значительные вопросы относительно последствий для создания, передачи и проверки знаний при использовании генеративного ИИ в контексте образования, преподавания и обучения, разработки и оценки учебных программ, исследований и авторского права. Большинство стран только начинают внедрять генеративный ИИ в образование, и долгосрочные последствия этого процесса еще предстоит осознать. Для обеспечения человеко-ориентированного использования ИИ необходимо срочно провести открытые общественные дебаты и политический диалог о долгосрочных последствиях. Инклюзивные обсуждения, в которых примут участие представители правительства, частного сектора и других заинтересованных сторон,

должны послужить основой для углубленного понимания и внесения вклада в итеративное обновление правил и политики.

3.3 Положение о генеративном ИИ: ключевые элементы

Всем странам необходимо должным образом регулировать генеративный ИИ, чтобы гарантировать, что он приносит пользу развитию образования и других контекстов. В этом разделе предлагаются действия по ключевым элементам, которые могут предпринять: (1) государственные регулирующие органы, (2) поставщики инструментов с поддержкой ИИ, (3) институциональные пользователи и (4) отдельные пользователи. Хотя многие элементы этой структуры носят транснациональный характер, все они также должны рассматриваться в свете местного контекста, то есть существующих образовательных систем конкретной страны и общих нормативных рамок.

3.3.1 Государственные регулирующие органы

Для координации разработки, согласования и реализации положений о генеративном ИИ необходим общеправительственный подход. Рекомендуются следующие семь ключевых элементов и действий:

- **Межсекторальная координация:** создать национальный орган, который будет руководить общегосударственным подходом к генеративному ИИ и координировать сотрудничество между секторами.
- **Согласование законодательства:** привести структуру в соответствие с соответствующими законодательными и нормативными условиями каждой страны – например, с общими законами о защите данных, положениями об интернет-безопасности, законами о безопасности данных, полученных от граждан или используемых для обслуживания граждан, и другими соответствующими законодательству и обычной практике. Необходимо оценить

уместность существующих правил и любые необходимые адаптации в ответ на новые проблемы, появившиеся из-за генеративного ИИ.

- **Баланс между регулированием генеративного ИИ и продвижением инноваций в области ИИ:** оказать содействие межсекторальному сотрудничеству между компаниями, организациями, образовательными и исследовательскими учреждениями, а также соответствующими государственными учреждениями для совместной разработки заслуживающих доверия моделей; поощрять создание экосистем с открытым исходным кодом для содействия совместному использованию ресурсов супер компьютера и высококачественных наборов данных для предварительного обучения; и способствовать практическому применению генеративного ИИ в различных секторах и созданию высококачественного контента на благо общества.
- **Оценка и классификация потенциальных рисков ИИ:** установить принципы и процесс оценки и категоризации эффективности, безопасности и защищенности услуг генеративного ИИ до их развертывания и на протяжении всего жизненного цикла системы. Необходимо рассмотреть механизмы категоризации, основанные на уровнях риска, который генеративный ИИ может представлять для граждан. Важно разделить их на строгие правила (т. е. запрет приложений или систем с поддержкой ИИ, несущих неприемлемый риск), специальные правила для приложений с высоким уровнем риска и общие правила для приложений, которые не относятся к категории высокого риска. Пример такого подхода представлен в проекте Закона ЕС об искусственном интеллекте.
- **Защита конфиденциальности данных:** принять во внимание тот факт, что использование генеративного ИИ почти всегда предполагает обмен пользователями своими данными с поставщиком генеративного ИИ. Необходимо уполномочить разработку и реализацию законов о защите личной информации пользователей, а также выявлять и бороться

с незаконным хранением, профилированием и распространением данных.

- **Определение и соблюдение возрастных ограничений для использования генеративного ИИ:** осознать, что большинство приложений генеративного ИИ в первую очередь предназначены для взрослых пользователей. Эти приложения часто влекут за собой существенные риски для детей, включая воздействие нежелательного контента, а также возможность манипулирования. В свете этих рисков и значительной неопределенности, которая по-прежнему окружает итеративные приложения генеративного ИИ, настоятельно рекомендуется использовать возрастные ограничения для технологий искусственного интеллекта общего назначения, чтобы защитить права и благополучие детей.

В настоящее время условия использования ChatGPT требуют, чтобы пользователям было не менее 13 лет, а пользователи младше 18 лет должны иметь разрешение своих родителей или законных опекунов на использование услуг.⁵² Эти возрастные ограничения или пороговые значения взяты из Закона США о защите конфиденциальности детей в интернете (Федеральная торговая комиссия, 1998 г.). Закон США, принятый в 1998 году до широкого распространения социальных сетей и задолго до создания простых в использовании и мощных приложений генеративного ИИ, таких как ChatGPT, определяет, что организациям или отдельным поставщикам социальных сетей не разрешается предоставлять услуги детям в возрасте до 13 лет без разрешения родителей. Многие граждане считают, что этот порог слишком низкий, и выступают за принятие закона о повышении возраста до 16 лет. В Общем регламенте по защите данных Европейского Союза (2016 г.) указано, что пользователям должно быть не менее 16 лет, чтобы пользоваться услугами социальных сетей без разрешения родителей.

Появление различных чат-ботов генеративного ИИ требует, чтобы страны тщательно продумали и публично

обсудили подходящий возрастной порог для независимого взаимодействия с платформами генеративного ИИ. Минимальный порог должен составлять 13 лет. Странам также необходимо будет решить, останется ли самостоятельное подтверждение возраста подходящим средством проверки возраста. Странам необходимо будет обязать поставщиков генеративного ИИ проверять возраст, а родителей или опекунов контролировать взаимодействие несовершеннолетних детей с инструментами генеративного ИИ.

- **Национальный контроль за данными и риск дефицита данных:** принятие законодательных мер для обеспечения защиты национальной собственности в информационной сфере и регулирование поставщиков генеративного ИИ, работающих в пределах государственных границ. Для наборов данных, созданных гражданами и используемых в коммерческих целях, установление правил, способствующих взаимовыгодному сотрудничеству, чтобы информация данной категории не покидала страну и не использовалась исключительно крупными информационными технологическими компаниями.

3.3.2. Поставщики инструментов генеративного ИИ

К поставщикам генеративного ИИ относятся организации и частные лица, которые отвечают за разработку и предоставление инструментов генеративного ИИ и/или используют технологии генеративного ИИ для предоставления услуг, в том числе через программируемые интерфейсы прикладного программирования (API). Большинство влиятельных поставщиков инструментов генеративного ИИ являются чрезвычайно хорошо финансируемыми компаниями. Поставщикам генеративного ИИ следует разъяснить, что они несут ответственность за соблюдение этических норм, в том числе за реализацию этических принципов, предусмотренных нормативными актами. Следует охватить следующие десять категорий ответственности:

- **Ответственность за человеческие ценности.** Поставщики генеративного ИИ должны нести ответственность за обеспечение соблюдения основных ценностей и законных целей, уважение интеллектуальной собственности и соблюдение этических норм, а также предотвращение распространения дезинформации и разжигания ненависти.
- **Надежные данные и модели.** Поставщики генеративного ИИ должны быть обязаны подтверждать надежность и этику источников данных и методов, используемых в их моделях и результатах. Они должны быть уполномочены принимать модели данных и основополагающих моделей с проверенными законными источниками и соблюдать соответствующие законы об интеллектуальной собственности (например, если данные защищены правами интеллектуальной собственности). Кроме того, когда моделям необходимо использовать личную информацию, сбор указанной информации должен происходить только с осознанного и явного согласия владельцев.
- **Создание недискриминационного контента.** Поставщики генеративного ИИ должны запретить разработку и внедрение систем генеративного ИИ, которые создают предвзятый или дискриминирующий контент на основе расы, национальности, пола или других характеристик. Им следует обеспечить наличие надежных «ограждений», предотвращающих создание оскорбительного, предвзятого или ложного контента, одновременно гарантируя, что люди, участвующие в информировании об «ограждениях», защищены.
- **Объясняемость и прозрачность моделей генеративного ИИ.** Поставщики обязаны предоставлять компетентным государственным органам разъяснения относительно источников, объема и характера данных, используемых в их моделях; их правила маркировки данных при предварительном обучении; методы или алгоритмы, которые их модели используют для генерации контента или ответов; и услуги, которые предоставляют их инструменты генеративного ИИ.

При необходимости они должны предложить поддержку, чтобы помочь государственным органам понять технологии и данные. Также пользователи должны быть осведомлены о возможности появления ошибок в сгенерированных ответах.

- **Маркировка контента генеративного ИИ.** В соответствии с соответствующими законами и положениями, касающимися синтеза онлайн-информации с помощью ИИ, поставщики должны маркировать документы, отчеты, изображения и видео, созданные генеративным ИИ, надлежащим и законным образом. Например, результат генеративного ИИ должен быть четко обозначен как контент, созданный машиной.
- **Принципы безопасности.** Поставщики генеративного ИИ должны обеспечивать безопасное, надежное и устойчивое обслуживание на протяжении всего жизненного цикла системы генеративного ИИ.
- **Спецификации доступа и использования.** Поставщики генеративного ИИ должны предоставить четкие спецификации в отношении соответствующей аудитории, сценариев и целей использования своих услуг, а также помогать пользователям инструментов генеративного ИИ принимать рациональные и ответственные решения.
- **Признание ограничений и предотвращение предсказуемых рисков.** Поставщики генеративного ИИ должны прозрачно информировать о том, какие методы используются системами и какие ограничения налагаются на их результаты. Важно, чтобы они разрабатывали технологии, гарантирующие, что входные данные, используемые методы и выходные данные не нанесут пользователю предсказуемого вреда, а также предусматривали протоколы для смягчения непредвиденного вреда в случае его возникновения. Кроме того, им следует предоставлять рекомендации, чтобы помочь пользователям осознанно воспринимать контент, созданный генеративным искусственным интеллектом на основе этических принципов, а также

предотвращать чрезмерную зависимость от такого контента.

- **Механизмы подачи жалоб и средства правовой защиты.** Поставщикам генеративного ИИ необходимо создать механизмы и каналы для сбора жалоб от пользователей и широкой общественности, а также принимать своевременные меры для принятия и обработки этих жалоб.
- **Мониторинг и отчетность о незаконном использовании.** Поставщики обязаны сотрудничать с государственными органами, чтобы облегчить контроль и предоставление отчетности о незаконном использовании контента, созданного генеративным ИИ. Это включает в себя ситуации, когда люди используют такой контент способами, которые противоречат закону или нарушают этические и социальные нормы, например распространение дезинформации, подстрекательство к ненависти, рассылка спама или создание вредоносного программного обеспечения.

3.3.3. Институциональные пользователи

К институциональным пользователям относятся органы образования и учреждения, такие как университеты и школы, которые несут ответственность за определение того, следует ли внедрять генеративный ИИ и какие типы инструментов генеративного ИИ следует внедрить в учреждениях.

- **Институциональный аудит алгоритмов, данных и результатов генеративного ИИ.** Внедрить механизмы для максимально эффективного мониторинга алгоритмов и данных, используемых инструментами генеративного ИИ, и результатов, которые они генерируют. Это должно включать регулярные проверки и оценки, защиту пользовательских данных и автоматическую фильтрацию нежелательного контента.
- **Проверка пропорциональности и защита благополучия пользователей.** Внедрить национальные механизмы классификации или создать институциональную политику для категоризации и проверки систем и приложений генеративного ИИ. Обеспечить, чтобы системы

генеративного ИИ, принятые учреждением, соответствовали утвержденным на местном уровне этическим нормам и не наносили предсказуемого вреда целевым пользователям учреждения, особенно детям и уязвимым группам.

- **Рассмотрение и устранение долгосрочных последствий.** Со временем использование инструментов или контента генеративного ИИ в образовании может оказать глубокое влияние на развитие человеческих способностей, таких как навыки критического мышления и творческие способности. Эти потенциальные последствия следует оценить и устранить.
- **Соответствие возрасту.** Рассмотреть возможность введения минимальных возрастных ограничений для самостоятельного использования генеративного ИИ в учреждении.

3.3.4 Индивидуальные пользователи

К индивидуальным пользователям потенциально относятся все люди во всем мире, имеющие доступ к интернету и хотя бы к одному типу инструментов генеративного ИИ. Используемый здесь термин «индивидуальные

пользователи» в основном относится к отдельным преподавателям, исследователям и учащимся формальных образовательных учреждений или тем, кто участвует в неформальных программах обучения.

- **Осведомленность о технических заданиях по использованию генеративного ИИ.** При подписании или одобрении условий соглашения об обслуживании пользователи обязаны быть в курсе своих обязательств по соблюдению технического задания, предусмотренных в соглашении, а также соответствовать законам или правилам, лежащим в основе данного соглашения.
- **Этичное использование приложений генеративного ИИ.** Пользователи должны ответственно подходить к использованию генеративного ИИ и избегать его использования способами, которые могут нанести ущерб репутации и законным правам других людей.
- **Мониторинг и сообщение о незаконных приложениях генеративного ИИ.** При обнаружении приложений генеративного ИИ, которые нарушают одно или несколько правил, пользователи должны уведомить государственные регулирующие органы.

4. На пути к созданию политической основы для использования генеративного ИИ в образовании и исследованиях

Регулирование генеративного ИИ с целью использования потенциальных преимуществ для образования и исследований требует разработки соответствующей политики. Данные опроса 2023 года, представленные выше, показывают, что лишь несколько стран приняли конкретные политики или планы по использованию искусственного интеллекта в образовании. В предыдущем разделе изложены концепция, необходимые шаги, а также ключевые элементы и действия, которые могут быть предприняты различными заинтересованными сторонами. В этом разделе представлены меры, которые можно предпринять для разработки последовательных, комплексных политических рамок для регулирования использования генеративного ИИ в образовании и исследованиях.

Отправной точкой для этого является «Технологии искусственного интеллекта в образовании: перспективы и последствия: руководство для лиц, ответственных за формирование политики» (ЮНЕСКО, 2022б). В нем предлагается комплексный набор рекомендаций, которые помогут правительствам разработать и реализовать общесекторальную политику в области ИИ и образования для оказания содействия качественному образованию, социальной справедливости и инклюзивности. Большинство рекомендаций остаются применимыми и могут быть в дальнейшем адаптированы для разработки конкретной политики по использованию генеративного ИИ в образовании. Следующие восемь конкретных мер по планированию политики генеративного ИИ в образовании и исследованиях предлагаются в дополнение к существующему руководству.

4.1 Содействие инклюзивности, равенству, языковому и культурному разнообразию

Важно осознавать, что инклюзивность играет критическую роль на протяжении всего существования и использования генеративного искусственного интеллекта. Особенно важно, чтобы инструменты генеративного ИИ способствовали решению основных проблем в сфере образования и соответствовали Целям устойчивого развития в области образования, а не создавали дополнительные барьеры или неравенства (например, на основе пола, этнической принадлежности, особенностей обучения, социально-экономического статуса, географического расположения, статуса мобильности и др.). Если данные инструменты не учитывают разнообразие и индивидуальные особенности, то они вредят равноправию, языковому разнообразию и культурному плюрализму. Для достижения данной цели рекомендуется принять три ключевые политические меры:

- Определить тех, кто не имеет или не может позволить себе подключение к интернету или данным, и принять меры по развитию универсального подключения и цифровых компетенций, чтобы уменьшить барьеры на пути к справедливому и инклюзивному доступу к инструментам ИИ. Создать устойчивые механизмы финансирования для разработки и предоставления инструментов с поддержкой искусственного интеллекта для учащихся с ограниченными возможностями или особыми потребностями. Продвигать использование генеративного ИИ для поддержки учащихся всех возрастов и с разным происхождением на протяжении всей жизни.

- Разработать критерии проверки систем генеративного ИИ, чтобы гарантировать отсутствие гендерных предубеждений, дискриминации в отношении маргинализированных групп или разжигания ненависти, заложенных в данные или алгоритмы.
- Разработать и внедрить инклюзивные спецификации для систем генеративного ИИ и реализовать институциональные меры для защиты языкового и культурного разнообразия при масштабном использовании генеративного ИИ в образовании и исследованиях. Соответствующие спецификации должны требовать от поставщиков генеративного ИИ включать данные на нескольких языках, особенно на местных или коренных языках, при обучении моделей GPT, чтобы улучшить способность генеративного ИИ реагировать на многоязычный текст и создавать его. Спецификации и институциональные меры должны строго запрещать поставщикам ИИ любое преднамеренное или непреднамеренное удаление языков меньшинств или дискриминацию в отношении носителей языков коренных народов, а также требовать от поставщиков удаления систем, продвигающих доминирующие языки или предвзятые культурные нормы.
- Проинформируйте учащихся о типах данных, которые генеративный ИИ может собирать, как эти данные используются и какое влияние они могут оказать на их образование и жизнь в целом.
- Поддерживайте внутреннюю мотивацию учащихся учиться и развиваться. Важно укрепить автономию людей в отношении их собственных подходов к исследованиям, преподаванию и обучению в контексте использования все более сложных систем генеративного ИИ.
- Ограничьте использование генеративного ИИ для того, чтобы предоставить учащимся возможности развивать когнитивные способности и социальные навыки путем наблюдения за реальным миром, экспериментов, обсуждений с другими людьми и путем независимых логических рассуждений.
- Обеспечьте достаточное социальное взаимодействие и соответствующее воздействие творческих объектов, создаваемых людьми, и не позволяйте учащимся становиться зависимыми от генеративного ИИ.
- Используйте инструменты генеративного ИИ, чтобы свести к минимуму нагрузку от выполнения домашних заданий и экзаменов, а не усугублять ее.

4.2 Защита человеческой свободы действий

Поскольку генеративный ИИ становится все более сложным, ключевой опасностью является его потенциал подорвать человеческую активность. Поскольку все больше людей используют генеративный ИИ для поддержки своей писательской или другой творческой деятельности, они могут непреднамеренно полагаться на него. Это может поставить под угрозу развитие интеллектуальных навыков. Хотя генеративный ИИ может быть использован для того, чтобы бросить вызов и расширить человеческое мышление, ему нельзя позволять поработить человеческое мышление. Защита и развитие человеческой деятельности всегда должны быть ключевыми опорами при разработке и внедрении генеративного ИИ с учетом следующих семи шагов:

- Проконсультируйтесь с исследователями, преподавателями и учащимися об их взглядах на генеративный ИИ и используйте обратную связь, чтобы решить, как следует использовать конкретные инструменты генеративного ИИ в институциональном масштабе. Поощряйте учащихся, преподавателей и исследователей критиковать и подвергать сомнению методологии, лежащие в основе систем искусственного интеллекта, точность генерируемого контента, а также нормы или педагогические подходы, которые они могут навязывать.
- Не допускайте передачи ответственности человека системам генеративного ИИ при принятии важных решений.

4.3 Мониторинг и проверка систем генеративного ИИ для образования

Как уже отмечалось, разработка и внедрение генеративного ИИ должны быть этичными. Впоследствии, как только генеративный ИИ будет использоваться, и на протяжении всего его жизненного цикла, его необходимо тщательно контролировать и проверять – на предмет его этических рисков, его педагогической целесообразности и соответствия, а также его влияния на учащихся, учителей и отношения в классе/школе. В связи с этим рекомендуются следующие пять действий:

- Создать механизмы контроля, чтобы проверить, свободны ли системы генеративного ИИ, используемые в образовании и исследованиях, от предвзятости, особенно гендерных предубеждений, и обучены ли они на данных, репрезентативных с точки зрения пола, возможностей здоровья, социального и экономического статуса, этнического и культурного происхождения, и географического положения.
- Решить сложную проблему информированного согласия, особенно в ситуациях, когда дети или другие уязвимые учащиеся не способны дать действительно информированное согласие.
- Проверить, включают ли результаты генеративного ИИ фейковые изображения, фейковые (неточные или ложные) новости, способствующие разжиганию ненависти. Если окажется, что генеративный ИИ создает неприемлемый контент, учреждения и преподаватели должны быть готовы и иметь возможность принять быстрые и решительные меры для уменьшения или устранения проблемы.
- Проводить строгую этическую проверку приложений генеративного ИИ, прежде чем они будут официально приняты в образовательных или исследовательских учреждениях (т. е. применять подход, основанный на этике).

- Прежде чем принимать решения об институциональном внедрении, важно убедиться, что рассматриваемые приложения генеративного ИИ не причиняют предсказуемого вреда учащимся, являются образовательными, эффективными и соответствуют возрасту и способностям учащихся. Они также должны отвечать здоровым педагогическим принципам (т. е. основаны на соответствующих областях), знаниям и ожидаемым результатам обучения.

4.4 Развитие у учащихся компетенций в области ИИ, включая навыки, связанные с генеративным ИИ

Развитие компетенций в области искусственного интеллекта среди учащихся играет важнейшую роль в обеспечении безопасного, этического и важного использования ИИ в сфере образования и за ее пределами. Тем не менее, по данным ЮНЕСКО, лишь около 15 стран к началу 2022 года разработали, внедрили или находились в процессе разработки утвержденных правительством учебных программ по ИИ в школах (ЮНЕСКО, 2022в). Последние достижения в области генеративного искусственного интеллекта еще более подчеркнули острую необходимость того, чтобы каждый человек достиг соответствующего уровня грамотности как в человеческом, так и в технологическом аспектах ИИ. Это включает в себя понимание его работы в широком смысле, а также изучение конкретного влияния генеративного ИИ. Для достижения этих целей на данный момент крайне важно предпринять следующие пять действий:

- Предоставить утвержденные правительством учебные программы по искусственному интеллекту для школьного образования, технического и профессионального образования и обучения, а также для обучения на протяжении всей жизни. Учебные программы по ИИ должны охватывать влияние ИИ на нашу жизнь, включая этические проблемы, понимание алгоритмов и данных, а также навыки

правильного и творческого использования инструментов ИИ, включая приложения генеративного ИИ;

- Поддерживать высшие учебные заведения и исследовательские учреждения в расширении программ развития местных талантов в области искусственного интеллекта;
- Продвигать гендерное равенство в развитии передовых компетенций в области искусственного интеллекта и создавать гендерно-сбалансированный коллектив профессионалов;
- Разрабатывать межсекторальные прогнозы национальных и глобальных изменений рабочих мест, вызванных автоматизацией, и повышать перспективные навыки на всех уровнях образования и систем непрерывного обучения на основе предполагаемых изменений;
- Обеспечить специальные программы для пожилых работников и граждан, которым, возможно, потребуется освоить новые навыки и адаптироваться к новым условиям.

4.5 **Наращивание потенциала преподавателей и исследователей для правильного использования генеративного ИИ**

Согласно данным опроса 2023 года об использовании ИИ в образовании правительствами (ЮНЕСКО, 2023*в*), только семь стран (Китай, Финляндия, Грузия, Катар, Испания, Таиланд и Турция) сообщили, что они разработали или разрабатывают системы обучения или обучающие программы по искусственному интеллекту для учителей. Только Министерство образования Сингапура сообщило о создании онлайн-репозитория, ориентированного на использование ChatGPT в преподавании и обучении. Это ясно показывает, что учителя в большинстве стран не имеют доступа к хорошо структурированному обучению по использованию ИИ и, в частности, генеративного ИИ в образовании.

Чтобы подготовить учителей к ответственному и эффективному использованию генеративного ИИ, странам необходимо предпринять следующие четыре действия:

- Сформулировать или скорректировать рекомендации на основе местных тестов, чтобы помочь исследователям и преподавателям ориентироваться в широкодоступных инструментах генеративного ИИ и управлять разработкой новых приложений ИИ для конкретной предметной области.
- Защитить права преподавателей и исследователей и ценность их практики при использовании генеративного ИИ. В частности, проанализировать уникальную роль учителей в содействии мышлению высшего порядка, организации человеческого взаимодействия и продвижении человеческих ценностей.
- Определить ценностную ориентацию, знания и навыки, необходимые учителям для понимания и эффективного и этичного использования систем генеративного ИИ. Предоставить учителям возможность создавать специальные инструменты на основе генеративного ИИ для содействия обучению и собственному профессиональному развитию.
- Провести динамический анализ компетенций, необходимых учителям для понимания и использования ИИ для преподавания, обучения и профессионального развития. Интегрировать новые наборы ценностей, знаний и навыков, связанных с ИИ, в структуры и программы компетенций для подготовки работающих и начинающих учителей.

4.6 **Продвижение плюрализма мнений и множественного выражения идей**

Как упоминалось ранее, генеративный ИИ не осознает представленные ответы. Вместо этого его ответы основаны на вероятностях языковых шаблонов, обнаруженных в данных (из интернета), которые были использованы

для его обучения. Для решения определенных фундаментальных проблем, связанных с его результатами, в настоящее время изучаются новые методы, такие как объединение генеративного ИИ с базами данных знаний и механизмами рассуждения. Тем не менее, из-за особенностей его функционирования, а также вследствие неучастливого отношения разработчиков, генеративный ИИ по своей природе воспроизводит доминирующие взгляды в своих результатах, подрывая при этом мнения меньшинства и плюрализма. Поэтому, если человеческая цивилизация стремится к процветанию, важно признать, что генеративный ИИ никогда не сможет быть авторитетным источником знаний по любой теме, которой он занимается.

В результате пользователям необходимо критически относиться к результатам, производимым генеративным ИИ. В частности:

- Осознать роль генеративного ИИ как быстрого, но зачастую ненадежного источника информации. Хотя некоторые плагины и инструменты на основе большой языковой модели, упомянутые ранее, предназначены для обеспечения доступа к проверенной и актуальной информации, на данный момент имеется мало убедительных доказательств того, что они эффективны.
- Поощрять учащихся и исследователей критически относиться к ответам, предоставленным генеративным ИИ. Важно признать, что генеративный ИИ обычно лишь повторяет устоявшиеся или стандартные мнения, тем самым подрывая множественность мнений и мнения меньшинства, а также многообразие выражения идей.
- Предоставить учащимся достаточные возможности учиться на основе проб и ошибок, эмпирических экспериментов и наблюдений за реальным миром.

4.7 Тестирование локально значимых прикладных моделей и формирование базы доказательств

В моделях генеративного искусственного интеллекта до сих пор преобладает информация с Глобального Севера, и недостаточно учитываются голоса Глобального Юга и коренных общин. Только благодаря решительным усилиям, например использованию синтетических данных (Марвала, 2023), инструменты генеративного ИИ станут более чувствительными к контексту и потребностям местных сообществ, особенно жителей стран Глобального Юга. Для изучения подходов, соответствующих местным потребностям, и для более широкого сотрудничества, рекомендуется выполнить следующие восемь действий:

- Гарантировать, что проектирование и внедрение генеративного ИИ планируются стратегически, а не происходят как часть пассивного и не очень важного процесса закупок.
- Поощрять дизайнеров генеративного ИИ ориентироваться на открытые, исследовательские и разнообразные варианты обучения.
- Тестировать и расширять научно обоснованные варианты применения ИИ в образовании и исследованиях, ориентируясь на образовательные приоритеты, а не на новинки, мифы или тренды.
- Направлять использование генеративного ИИ для стимулирования инноваций в исследованиях, в том числе за счет использования вычислительных возможностей, крупномасштабных данных и результатов генеративного ИИ для информирования и стимулирования совершенствования исследовательских методологий.
- Проводить оценку социальных и этических последствий включения генеративного ИИ в исследовательские процессы.

- Для создания доказательной базы эффективности генеративного искусственного интеллекта в контексте поддержки инклюзивного образования, достижения учебных целей, и исследований, а также для продвижения языкового и культурного разнообразия, необходимо установить конкретные критерии, опираясь на подтвержденные педагогические исследования и методологии.
- Проводить последовательные мероприятия по улучшению фактической информации о социальном и этическом воздействии генеративного искусственного интеллекта.
- Проанализировать экологические издержки, связанные с широкомасштабным использованием технологий искусственного интеллекта, включая энергозатраты и ресурсы, необходимые для обучения моделей, таких как GPT, и разработать устойчивые цели. Цели должны быть достигнуты поставщиками искусственного интеллекта с целью предотвращения увеличения негативного воздействия на климат.
- снизить риски, связанные с генеративным искусственным интеллектом в образовании и научных исследованиях.
- Объединить межсекторальный и междисциплинарный опыт преподавателей, исследователей, студентов-ученых, специалистов в области ИИ и представителей других заинтересованных сторон для динамического изучения долгосрочных последствий использования генеративного ИИ в образовании, научно-исследовательской деятельности, авторском праве, учебных программах и методах оценки, а также в человеческом взаимодействии и социальных сетях.
- Предоставлять своевременные рекомендации для информирования о регулярных обновлениях правил и политик.

4.8 Анализ долгосрочных последствий на межсекторальной и междисциплинарной основе

Межсекторальные и междисциплинарные подходы необходимы для эффективного и этичного использования генеративного ИИ в образовании и исследованиях. Только благодаря обширным экспертным знаниям и объединению множества заинтересованных сторон мы сможем быстро выявить и успешно решить основные проблемы, чтобы уменьшить долгосрочные негативные последствия, и при этом в полной мере воспользоваться преимуществами. В этой связи рекомендуется выполнить следующие три действия:

- Взаимодействовать с поставщиками искусственного интеллекта, преподавателями, исследователями, родителями и учащимися для разработки системных корректировок в учебных программах и методах оценки, с тем чтобы в полной мере использовать потенциал и

5. Содействие творческому использованию генеративного ИИ в образовании и исследованиях

Когда ChatGPT впервые был запущен, преподаватели по всему миру выразили обеспокоенность относительно его способности генерировать эссе и потенциальной возможности помогать учащимся в списывании. Недавно многие люди и организации, включая несколько ведущих университетов мира, заявили, что «джин вышел из бутылки», и такие инструменты, как ChatGPT, не только продолжат существование, но могут быть продуктивно использованы в образовательных учреждениях. В настоящее время интернет полон предложений по использованию генеративного ИИ в образовании и научных исследованиях. Это включает его применение для поиска новых идей, создания многоаспектных примеров, разработки учебных планов и презентаций, обобщения существующего материала и стимулирования создания изображений. Несмотря на появление новых идей практически каждый день в онлайн-среде, исследователи и преподаватели по-прежнему изучают влияние генеративного ИИ на образование, обучение и исследования. В частности, те, кто предлагает различные варианты использования, возможно, недостаточно учитывали этические принципы, в то время как другие ориентировались на технологический потенциал генеративного ИИ, а не на потребности исследователей, преподавателей или учащихся. В данном разделе описываются пути поощрения творческого использования генеративного ИИ в образовании.

5.1 Институциональные стратегии, способствующие ответственному и творческому использованию генеративного ИИ

Как упоминалось ранее, образовательные и исследовательские учреждения должны разрабатывать, внедрять и проверять соответствующие стратегии и этические нормы

для руководства ответственным и этичным использованием систем и приложений генеративного ИИ, с тем чтобы удовлетворять потребности в обучении, преподавании и научных исследованиях. Данный результат можно достигнуть путем использования следующих четырех стратегий:

- **Институциональная реализация этических принципов:** обеспечить, чтобы исследователи, преподаватели и учащиеся использовали инструменты генеративного ИИ ответственно и этично, а также критически подходили к точности и достоверности результатов.
- **Руководство и обучение:** предоставить исследователям, преподавателям и учащимся рекомендации и обучение по использованию инструментов генеративного ИИ, чтобы убедиться, что они понимают этические проблемы, такие как предвзятость в маркировке данных и алгоритмах, и что они соблюдают соответствующие правила конфиденциальности данных и интеллектуальной собственности.
- **Создание потенциала в области оперативного проектирования генеративного ИИ.** Помимо знаний по конкретным предметам, исследователям и преподавателям также понадобятся знания в области разработки и критической оценки результатов, воспроизводимых генеративным ИИ. В связи с тем, что проблемы, возникшие в связи с генеративным искусственным интеллектом, являются сложными, необходимо обеспечить исследователям и преподавателям качественную подготовку и поддержку.
- **Обнаружение плагиата в письменных заданиях.** Генеративный ИИ может позволить учащимся выдавать текст, который они не писали, за свою собственную работу — новый тип

«плагиата». Поставщики генеративного ИИ должны маркировать свою продукцию водяными знаками «создано при помощи ИИ», в то же время разрабатывая средства для идентификации материалов, созданных с использованием ИИ. Однако имеется ограниченное количество доказательств эффективности данных мер и средств. Прямой институциональной стратегией является поддержка академической честности и усиление ответственности путем тщательного мониторинга со стороны людей. Долгосрочной стратегией является пересмотр структуры письменных заданий, чтобы их использование не приводило к оценке задач, которые могут быть выполнены инструментами генеративного ИИ лучше, чем учащимися. Вместо этого, преподаватели должны сосредоточиться на возможностях, которые доступны только людям, и недоступны генеративному ИИ и другим средствам ИИ. Это включает применение человеческих ценностей, таких как сострадание и творчество, в решении сложных проблем реального мира.

5.2 Подход, ориентированный на человека и педагогически целесообразное взаимодействие

Исследователи и преподаватели должны уделять приоритетное внимание человеческой деятельности и ответственному, педагогически обоснованному взаимодействию между людьми и инструментами ИИ при принятии решения о том, следует ли и как использовать генеративный ИИ. Это включает пять аспектов:

- использование инструмента(ов) должно способствовать удовлетворению потребностей людей, делая процессы обучения или исследований более эффективными по сравнению с нетехнологическим или альтернативным подходом;
- использование инструмента(ов) преподавателями и учащимися должно основываться на их внутренней мотивации;
- процесс использования инструмента(ов) должен контролироваться

преподавателями, учащимися или исследователями;

- выбор и организация инструментов, а также создаваемого ими контента, должны соответствовать возрасту учащихся, ожидаемым результатам и типу целевых знаний (например, фактических, концептуальных, процедурных или метакогнитивных), а также целевым проблемам (например, хорошо структурированным или плохо структурированным).
- Процессы использования должны способствовать интерактивному взаимодействию людей с генеративным ИИ, способствовать развитию мышления высокого порядка, а также нести ответственность за принятие решений в отношении точности создаваемого ИИ контента, стратегий обучения или исследований, а также их влияния на поведение человека.

5.3 Совместная разработка использования генеративного ИИ в образовании и исследованиях

Применение генеративного ИИ в сфере образования и научных исследований не должно быть ни установлено сверху вниз, ни мотивировано коммерческой пользой. Вместо этого безопасное и эффективное использование должно быть разработано совместно преподавателями, учащимися и исследователями. Также требуется надежный процесс пилотирования и оценки, чтобы изучить эффективность и долгосрочное воздействие различных видов применения генеративного ИИ.

Чтобы облегчить предлагаемую совместную разработку, в данном руководстве предлагается структура, включающая следующие шесть принципов, чтобы объединить педагогически оправданные практики и определить приоритеты человеческой деятельности:

- соответствующие области знаний или проблем;

- ожидаемые результаты;
- соответствующие инструменты генеративного ИИ и сравнительные преимущества;
- требования к пользователям;
- необходимые педагогические методы, созданные человеком и образцы запросов генеративному ИИ;
- этические риски.

В данном разделе приводятся примеры того, как процесс совместного проектирования при использовании генеративного ИИ может влиять на исследовательскую практику, способствовать преподаванию, обеспечивать наставничество для самостоятельного приобретения базовых навыков, способствовать развитию более высоких форм мышления и поддерживать учащихся с особыми потребностями. Эти примеры

представляют лишь малую часть обширного спектра областей, в которых генеративный ИИ может приносить пользу.

5.3.1 Генеративный ИИ для исследований

Модели генеративного ИИ показали свой потенциал в расширении обзора исследовательских направлений, обогащении анализа данных и рассмотрении литературных обзоров (см. **Таблицу 3**). Несмотря на то, что представлено более широкое разнообразие вариантов использования, требуются новые исследования для выявления потенциальных областей исследовательских задач и ожидаемых результатов, демонстрации эффективности и точности, а также обеспечения того, что человеческая деятельность при понимании реального мира через исследования не будет подвергнута риску из-за использования ИИ.

Таблица 3. Совместная разработка использования генеративного ИИ в научных исследованиях

Потенциальное, но недоказанное использование	Соответствующие области знаний или проблем	Ожидаемые результаты	Соответствующие инструменты генеративного ИИ и сравнительные преимущества	Требования к пользователям	Необходимые педагогические методы, созданные человеком, и примеры запросов генеративному ИИ	Возможные риски
Искусственный интеллект для подготовки структуры научного исследования	Генеративный ИИ может оказаться полезным в хорошо структурированных областях исследовательских задач.	Разработка и подготовка ответов на исследовательские вопросы, предложение соответствующих методологий. Потенциальная трансформация: индивидуальный подход при планировании исследований	Необходимо оценить начиная со списка в разделе 1.2, являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Исследователь должен иметь базовое понимание тем(ы). Исследователь должен развивать способность проверять информацию и уметь определять цитаты из несуществующих научных статей.	Основные идеи для определения проблем исследования (например, целевая аудитория, исследовательские вопросы, контекст), а также методологии, ожидаемые результаты и форматы. Пример запроса: <i>Предложите 10 потенциальных исследовательских вопросов по [теме x] и ранжируйте их по важности для [области исследования].</i>	Необходимо обратить внимание на высокий риск возможности создания генеративным ИИ ложной информации, например предоставление несуществующих исследовательских публикаций. Также существует риск того, что у исследователей появится соблазн копировать и вставлять схемы исследований, созданные ИИ. Это может ограничить возможности молодых исследователей для обучения методом проб и ошибок.

Потенциальное, но недоказанное использование	Соответствующие области знаний или проблем	Ожидаемые результаты	Соответствующие инструменты генеративного ИИ и сравнительные преимущества	Требования к пользователям	Необходимые педагогические методы, созданные человеком, и примеры запросов генеративному ИИ	Возможные риски
Генеративный ИИ для сбора данных и обзора литературы	Генеративный ИИ может оказаться полезным в плохо структурированных областях исследовательских задач	Автоматизированный сбор информации, исследование обширного массива данных, предоставление литературных источников и автоматизация части процесса интерпретации данных Потенциальная трансформация: ИИ-тренер для изучения данных и обзора литературы	Необходимо оценить начиная со списка в разделе 1.2, являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Исследователи должны обладать глубокими знаниями методологий и методов анализа данных.	Прогрессивное определение проблем, объема данных и литературных источников, используемых для проведения исследования данных и обзоров литературы, применяемых методологий, а также ожидаемых результатов и их предполагаемых форматов.	Необходимо проявлять осторожность в отношении информации, предоставляемой генеративным ИИ, избегать неправомерной обработки данных, возможных нарушений конфиденциальности, несанкционированного профилирования и гендерной дискриминации. Также важно быть внимательным к распространению доминирующих норм и их потенциальной угрозе для альтернативных норм и разнообразия мнений.

5.3.2 Генеративный ИИ для поддержки обучения

Использование как общих платформ генеративного искусственного интеллекта, так и конкретных образовательных инструментов должно быть направлено на улучшение понимания учителями своих предметов, а также их знаний о методологиях преподавания. Это может быть достигнуто через совместное проектирование учителями и искусственным интеллектом планов уроков, курсов или целых учебных программ. ИИ-помощники

для учителей или «генеративные близнецы учителей»,⁵³ которые предварительно обучены на основе данных опытных учителей и библиотек, прошли тестирование в некоторых учебных заведениях и имеют неизведанный потенциал, а также неосвоенные этические риски. Процессы практического применения и последующие итерации этих моделей продолжают требовать внимательного аудита в соответствии со структурой, рекомендованной в данном руководстве, а также обеспечения контроля со стороны человека, что отражено в **Таблице 4**.

Таблица 4. Совместная разработка использования генеративного ИИ для поддержки учителей и обучения

Потенциальное, но недоказанное использование	Соответствующие области знаний или проблем	Ожидаемые результаты	Соответствующие инструменты генеративного ИИ и сравнительные преимущества	Требования к пользователям	Необходимые педагогические методы, созданные человеком, и примеры запросов генеративному ИИ	Возможные риски
Генеративный ИИ в роли соавтора учебной программы или курса.	Знания концептуального характера по определенным предметам преподавания и процедурные знания методологий преподавания.	Помощь в разработке учебной программы и уроков, включая изложение или расширение концепций в ключевых областях целевой темы и формирование структуры учебного материала. Также возможна поддержка учителей при подготовке тестов и экзаменов через предоставление примеров вопросов и критериев для оценки. Потенциальная трансформация: учебная программа, созданная с помощью ИИ	Необходимо оценить начиная со списка в разделе 1.2, являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Учителя должны понимать и тщательно определять, чего они хотят достичь в рамках учебной программы, курсов, уроков или тестов, хотя бы они обратились к процедурным или концептуальным знаниям и какую теорию преподавания они хотят применить.	Вопросы к генеративному ИИ касаются предложения структуры и предоставления фактических знаний по теме(ам), а также выработки методов и процессов преподавания по различным вопросам или темам. Также рассматривается возможность создания пакетов курсов или планов уроков на основе конкретных тем и форматирования. Разработчикам учебных программ следует проводить проверку фактических знаний и удостоверяться в соответствии предлагаемых пакетов курсов.	Существует высокий риск того, что генеративный ИИ навязывает доминирующие нормы и педагогические методы. Это может непреднамеренно укрепить практику исключения в пользу уже обладающих обширными данными групп и усилить неравенства в доступе к соответствующим и высококачественным образовательным возможностям, создавая невыгодное положение для групп с ограниченным объемом данных.
Генеративный чат-бот в качестве помощника учителя	Концептуальные знания в нескольких областях в хорошо структурированных проблемах.	Предоставление индивидуальной поддержки, ответы на вопросы и поиск ресурсов. Потенциальная трансформация: генеративный близнец учителя	Необходимо оценить начиная со списка в разделе 1.2, являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Поддерживая учителей, данная система напрямую ориентирована на обучающихся, что предполагает наличие у них достаточных предварительных знаний, навыков и метакогнитивных умений для проверки результатов генеративного ИИ и выявления дезинформации. Таким образом, эта система может быть более подходящей для учащихся высших учебных заведений.	Требует от учителей ясного понимания проблем, контроля за ходом обсуждения и помощи учащимся в проверке сомнительных ответов, предоставленных генеративным ИИ.	Основываясь на текущих возможностях моделей генеративного ИИ, образовательные учреждения должны обеспечивать человеческий контроль за ответами, предоставляемыми инструментами генеративного ИИ, с учетом риска дезинформации. Это также может привести к ограничению доступа учащихся к человеческому руководству и поддержке, что затрудняет развитие прочных отношений между учителем и учеником, что вызывает особые опасения в отношении учащихся.

5.3.3 Генеративный ИИ в качестве индивидуального наставника для самостоятельного освоения базовых навыков

Несмотря на растущее внимание к развитию мышления высшего порядка и креативности при определении результатов обучения, важность базовых навыков для психологического развития детей и формирования компетентности не подлежит сомнению. Среди разнообразия умений эти фундаментальные навыки включают аудирование, произношение и письмо на

родном или иностранном языке, а также основы счета, рисования и программирования. Подход «упражнение и практика» не следует рассматривать как устаревший педагогический метод; вместо этого его стоит активно использовать и модернизировать с применением технологий генеративного ИИ, чтобы содействовать самостоятельному освоению учащимися базовых навыков. При соблюдении этических и педагогических принципов инструменты генеративного ИИ могут стать индивидуальными тренерами для практики самостоятельного обучения, как показано в **Таблице 5**.

Таблица 5. Совместная разработка использования генеративного ИИ в качестве индивидуального наставника для самостоятельного освоения базовых навыков в языках и искусстве

Потенциальное, но недоказанное использование	Соответствующие области знаний или проблем	Ожидаемые результаты	Соответствующие инструменты генеративного ИИ и сравнительные преимущества	Требования к пользователям	Необходимые педагогические методы, созданные человеком, и примеры запросов генеративному ИИ	Возможные риски
Индивидуальный тренер по языковым навыкам	Изучение языка, включая разговорную практику.	Активизация разговорной практики с целью улучшения навыков аудирования, устной и письменной речи, предоставление обратной связи, коррекции и моделирование родного или иностранного языка. Помощь учащимся в совершенствовании навыков письма. Потенциальная трансформация: Индивидуальные уроки по языку для начального уровня	Необходимо оценить начиная со списка в разделе 1.2, являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Необходимо установить возрастные ограничения, учитывая неприемлимый или несоответствующий возрасту контент, предоставляемый системами генеративного ИИ. Учащийся должен обладать внутренней мотивацией для участия в разговоре с системой ИИ. Учащийся должен быть способен критически оценивать предложения, представленные генеративным ИИ и проверять их точность.	При использовании общих платформ генеративного ИИ преподаватели могут направлять учащихся к использованию инструментов генеративного ИИ для получения обратной связи с целью улучшения, коррекции произношения или предоставления примеров письма. Например: <i>Вступите со мной в разговор на языке [x], помогая мне непрерывно совершенствоваться.</i> <i>Предложите несколько идей, которые помогут мне написать о [теме x].</i>	Необходимо проявлять внимание к культурно нечувствительным или неточным выражениям, а также избегать непреднамеренного закрепления стереотипов или культурных предубеждений. Без применения соответствующих педагогических стратегий, стимулирующих внутреннюю мотивацию учащихся, это может ограничить творческий потенциал и оригинальность детей, что приведет к шаблонному подходу к письму. Это также может ограничивать возможности для реального взаимодействия, разнообразия мнений, свободного выражения и критического мышления.

5. Содействие творческому использованию генеративного ИИ в образовании и исследованиях

Потенциальное, но недоказанное использование	Соответствующие области знаний или проблем	Ожидаемые результаты	Соответствующие инструменты генеративного ИИ и сравнительные преимущества	Требования к пользователям	Необходимые педагогические методы, созданные человеком, и примеры запросов генеративному ИИ	Возможные риски
Индивидуальный художественный тренер	Технические навыки в таких областях искусства, как музыка и рисование.	Предоставление предложений по художественным приемам (например, советы по перспективе и цвету) или музыкальной композиции (например, по мелодии и последовательности аккордов). Потенциальная трансформация: Индивидуальный преподаватель рисования для начального уровня	Необходимо оценить начиная со списка в разделе 1.2, являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Учащиеся должны предварительно определить свои цели в создании искусства или музыки, обладать базовым пониманием ключевых элементов в соответствующей области искусства или музыки, а также иметь базовые навыки анализа произведений искусства или музыкальных композиций.	Преподаватели должны предложить учащимся сравнивать художественные приемы ИИ с их собственными произведениями искусства. Учителя или тренеры должны поощрять учащихся развивать и применять свое воображение и творческие способности, которые генеративный ИИ не может заменить. Пример запроса: <i>Предложите несколько идей, которые вдохновят меня на создание изображения по [темам/идеям].</i>	Инструменты генеративного ИИ могут представлять учащимся неприемлемый или оскорбительный контент, что может нарушить их право на безопасность и благополучие. Инструменты генеративного ИИ увеличивают риск того, что учащиеся утратят способность развивать свое воображение и творческие способности.
Индивидуальный тренер по кодированию или арифметике	Навыки и знания в области концептуального программирования на начальном уровне. Это также может охватывать изучение основ математики.	Поддержка самостоятельного обучения основам кодирования, выявление ошибок в коде учащихся и предоставление мгновенной обратной связи, а также адаптация ответов на вопросы. Потенциальная трансформация: Индивидуальный преподаватель кодирования на начальном уровне	Необходимо оценить начиная со списка в разделе 1.3, являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Поиск и определение проблемы, а также разработка алгоритмов для ее решения остаются основными аспектами обучения кодированию и программированию. Учащиеся должны обладать внутренней мотивацией для использования программирования, а также некоторыми базовыми знаниями и навыками работы с языком программирования.	Учителя и тренеры должны обучать основам знаний и навыков, а также мотивировать учащихся использовать вычислительное мышление и программирование для решения проблем, включая совместное создание кода. Пример запроса: <i>Предложите несколько необычных идей по кодированию.</i>	Точность обратной связи и предложений остаётся проблематичным вопросом, поскольку генеративный ИИ не всегда будет прав. Существует высокий риск того, что инструменты генеративного ИИ могут мешать развитию у учащихся навыков вычислительного мышления и способностей выявлять и определять значимые проблемы для кодирования.

5.3.4 Генеративный ИИ для содействия исследованиям или проектному обучению

Если инструменты генеративного ИИ не используются специально для поддержки более высоких уровней мышления или творчества, они могут стимулировать плагиат или поверхностное «стохастическое повторение» результатов. Тем не менее,

учитывая, что модели генеративного ИИ были обучены на обширных данных, у них есть потенциал выступить в качестве оппонента в сократических диалогах или помощника исследователя в проектном обучении. Однако этот потенциал может быть реализован только через процессы разработки учебных планов, направленные на стимулирование более высоких уровней мышления, как показано в **Таблице 6**.

Таблица 6. Совместная разработка использования генеративного ИИ для содействия исследованиям или проектному обучению

Потенциальное, но недоказанное использование	Соответствующие области знаний или проблем	Ожидаемые результаты	Соответствующие инструменты генеративного ИИ и сравнительные преимущества	Требования к пользователям	Необходимые педагогические методы, созданные человеком, и примеры запросов генеративному ИИ	Возможные риски
Генеративный ИИ в роли оппонента в сократическом диалоге	Плохо структурированные проблемы.	Учащиеся вовлечены в диалог, напоминающий сократовское сомнение в предшествующих знаниях, что приводит к открытию новых знаний или более глубокому пониманию. Потенциальная трансформация: Индивидуальный оппонент в сократическом диалоге	Необходимо оценить начиная со списка в разделе 1.3, являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Учащийся должен достичь того возраста, который позволяет ему вести независимые беседы с инструментами генеративного ИИ. Учащиеся должны обладать предварительными знаниями и навыками, чтобы проверять точность представленных аргументов и информации.	Преподаватели могут подготовить список вопросов для обогащения диалога в качестве примеров для учащихся. Учащиеся также могут начать с широкого запроса, например: «Создайте сократический диалог, чтобы помочь мне критически взглянуть на [тему x]», и затем постепенно обогащать диалог более конкретными вопросами.	Существующие инструменты генеративного ИИ могут предоставлять схожие или стандартные ответы, что ограничивает доступ учащихся к различным точкам зрения и альтернативным перспективам, вызывая эффект эхо-камеры и затрудняя развитие независимого мышления.

5. Содействие творческому использованию генеративного ИИ в образовании и исследованиях

Потенциальное, но недоказанное использование	Соответствующие области знаний или проблем	Ожидаемые результаты	Соответствующие инструменты генеративного ИИ и сравнительные преимущества	Требования к пользователям	Необходимые педагогические методы, созданные человеком, и примеры запросов генеративному ИИ	Возможные риски
Генеративный ИИ в роли консультанта по проектному обучению	Плохо структурированные исследовательские проблемы в науке или социальных исследованиях.	Содействие формированию знаний, предоставляя поддержку учащимся в организации обучения на основе проектов. Это включает в себя функцию генеративного искусственного интеллекта, аналогичную роли научного руководителя, описанной в Таблице 3. Потенциальная трансформация: Индивидуальный тренер по проектному обучению	Необходимо оценить начиная со списка в разделе 1.3, являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Учащиеся могут выступать в роли младших исследователей при планировании и реализации обучения на основе проектов. Они должны быть достаточно зрелыми для самостоятельного использования платформ генеративного ИИ. Учащиеся должны обладать мотивацией и способностью участвовать в самостоятельной учебной деятельности на основе проектов, чтобы предотвращать возможность пассивного копирования и вставки ответов, предоставляемых инструментами генеративного ИИ.	Учителя направляют учащихся обращаться к генеративному ИИ для получения основных идей при определении проблем исследования, как указано в разделе 5.3.1. Учащиеся, работая индивидуально или в группах, используют инструменты генеративного ИИ для проведения обзоров литературы, сбора и обработки данных, а также создания отчетов.	Учащиеся, не обладающие глубокими предварительными знаниями и способностью, необходимой для проверки точности ответов, могут быть введены в заблуждение информацией, предоставляемой инструментами генеративного ИИ. Это также может ограничить взаимодействие учащихся со сверстниками, сократив возможности для совместного обучения и потенциально нанося вред их социальному развитию.

5.3.5 Генеративный ИИ для поддержки учащихся с особыми потребностями

Теоретически модели генеративного ИИ могут помочь учащимся с нарушениями слуха или зрения. Новые практики включают субтитры или субтитры с поддержкой генеративного ИИ для глухих и слабослышащих учащихся, а также аудио описание, созданное генеративным ИИ, для учащихся с нарушениями зрения. Модели генеративного ИИ также могут преобразовывать текст в речь и речь в текст, чтобы люди с нарушениями зрения, слуха или речи могли получать доступ к контенту, задавать вопросы и общаться со своими сверстниками. Однако эта функция еще не получила широкого применения. Согласно упомянутому ранее опросу об использовании правительствами ИИ в образовании,

проведенному ЮНЕСКО в 2023 году, только четыре страны (Китай, Иордания, Малайзия и Катар) сообщили, что их правительственные учреждения проверили и рекомендовали инструменты с использованием ИИ для поддержки инклюзивного доступа для учащихся с ограниченными возможностями (ЮНЕСКО, 2023в).

Существует также тенденция к обучению итераций моделей генеративного ИИ, чтобы помочь учащимся использовать родные языки, включая языки меньшинств и коренных народов, для обучения и общения. Например, PaLM 2, большая языковая модель нового поколения от Google, обучается на параллельных данных, охватывающих сотни языков в виде пар исходного и целевого текста. Включение параллельных многоязычных

данных призвано еще больше улучшить способность модели понимать и генерировать многоязычный текст (Google, 2023б).

Предоставляя переводы, перефразирование и автоматическое исправление в режиме реального времени, инструменты генеративного ИИ могут помочь учащимся, использующим языки меньшинств, передавать идеи и улучшать сотрудничество со сверстниками из разных языковых слоев. Однако это не произойдет естественным путем в таком масштабе. Только при целенаправленном планировании можно

реализовать потенциал для усиления голоса маргинализированных групп.

Наконец, также высказывалась гипотеза о том, что системы генеративного ИИ обладают потенциалом для проведения диагностики на основе разговора, выявления психологических или социально-эмоциональных проблем, а также трудностей в обучении. Однако до сих пор мало доказательств того, что этот подход эффективен или безопасен, и любой диагноз потребует интерпретации квалифицированными специалистами.

Таблица 7. Совместная разработка использования генеративного ИИ для поддержки учащихся с особыми потребностями

Потенциальное, но недоказанное использование	Соответствующие области знаний или проблем	Ожидаемые результаты	Соответствующие инструменты генеративного ИИ и сравнительные преимущества	Требования к пользователям	Необходимые педагогические методы, созданные человеком, и примеры запросов генеративному ИИ	Возможные риски
Консультативная диагностика с целью выявления трудностей в обучении	Этот способ может быть весьма полезным для учащихся, сталкивающихся с трудностями в обучении, обусловленными психологическими, социальными или эмоциональными проблемами.	Применение естественного языка для выявления потребностей учащихся, сталкивающихся с психологическими, социальными или эмоциональными трудностями в обучении, с целью предоставления им соответствующей поддержки или обучения. Потенциальная трансформация: Основной консультант по вопросам обучения и социально-эмоциональных трудностей учащихся.	Дополнительно к общим инструментам генеративного ИИ, нужно найти чат-боты, основанные на генеративных моделях ИИ. Необходимо оценить являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями. Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.	Учителям или специалистам, взаимодействующим с этой группой учащихся, важно удостовериться в точности основных рекомендаций, предоставляемых генеративным ИИ.	Учителя или координаторы должны создать комфортные условия для разговора с учащимся, с целью выявления психологических, социальных или эмоциональных проблем, а также трудностей в обучении.	Генеративный ИИ может ошибочно определить конкретные трудности учащегося, что влечет за собой возможность предоставления несоответствующей поддержки.

Потенциальное, но недоказанное использование	Соответствующие области знаний или проблем	Ожидаемые результаты	Соответствующие инструменты генеративного ИИ и сравнительные преимущества	Требования к пользователям	Необходимые педагогические методы, созданные человеком, и примеры запросов генеративному ИИ	Возможные риски
Инструменты обеспечения доступности на основе ИИ	Инструменты обеспечивают учащимся с нарушениями слуха или зрения более широкий доступ к ресурсам, что в свою очередь повышает качество их обучения.	<p>Обеспечение удовлетворения потребностей учащихся в доступе к знаниям по конкретным предметам осуществляется путем предоставления субтитров с использованием генеративного ИИ и/или сурдоперевода для аудио- или видеоконтента, а также создания аудио описаний для текста или других визуальных материалов.</p> <p>Потенциальная трансформация: Персонализированная языковая поддержка, основанная на ИИ</p>	<p>В дополнение к стандартным инструментам генеративного ИИ нужно найти соответствующие и надежные решения для создания субтитров и аудио описаний, использующие ИИ.</p> <p>Необходимо оценить являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями.</p> <p>Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.</p>	Учителя и координаторы должны помочь учащимся в освоении инструментов генеративного ИИ, обеспечивая доступ к ним. Важно также удостовериться, что результаты использования этих инструментов действительно поддерживают учащихся и не углубляют сложности и предвзятости, с которыми они сталкиваются.	<p>Необходимо предварительно проверить доступность платформ и инструментов, чтобы выявить и устранить возможные проблемы перед их использованием. Инструменты генеративного ИИ предоставляют доступ к ресурсам, поэтому учителям следует сосредоточиться на повышении качества обучения и социального благополучия.</p> <p>Учителя и координаторы должны обучить учащихся создавать голосовые или текстовые запросы для генеративного искусственного интеллекта в зависимости от их способностей.</p>	<p>Субтитры или аудио описания, создаваемые платформами генеративного ИИ, не предназначенными специально для поддержки зрения или слуха, часто бывают неточными и могут ввести в заблуждение учащихся с особыми потребностями.</p> <p>Эти инструменты могут непреднамеренно усилить существующие предубеждения.</p>
Генеративный ИИ для повышения качества образования учащихся, находящихся в маргинальном положении.	Для учащихся, принадлежащих к языковому или культурному меньшинству, активное участие в онлайн-пространстве и проведение совместных социальных исследований могут эффективно способствовать выражению и усилению их голоса.	<p>Предоставление переводов, перефразирования и автоматической коррекции текста в реальном времени для помощи учащимся из маргинализованных групп в использовании собственного языка в общении со сверстниками из различных языковых групп.</p> <p>Потенциальная трансформация: Инклюзивные языковые модели для обучения учащихся из маргинализованных групп.</p>	<p>Примером для рассмотрения является языковая модель PaLM 2.</p> <p>Необходимо оценить являются ли инструменты генеративного ИИ локально доступными, с открытым исходным кодом, тщательно протестированы или подтверждены властями.</p> <p>Далее следует рассмотреть преимущества и проблемы каждого конкретного инструмента генеративного ИИ и убедиться, что он должным образом отвечает определенным потребностям человека.</p>	Учащиеся должны обладать соответствующими знаниями или значимыми мнениями по теме обсуждения или совместного исследования. Они должны быть способными вносить ответственный и недискриминационный вклад, избегая разжигания ненависти.	Учителя или преподаватели должны разрабатывать для учащихся исследования и письменные задания на социальные или культурные темы или организовывать онлайн-семинары или межкультурное взаимодействие с целью поощрения учащихся в разработке идей и обмена мнениями.	<p>Важно выявлять и корректировать ошибки в переводах и переформулировках ИИ, которые могут привести к межкультурному недопониманию.</p> <p>Несмотря на то, что такое применение может предоставить маргинализированным учащимся возможность усилить свой голос, оно не затрагивает основные причины недостатка данных.</p>

6. Генеративный ИИ и будущее образования и исследований

Технологии генеративного искусственного интеллекта продолжают динамичное развитие и, вероятно, окажут значительное воздействие на области образования и исследований. Поэтому, требуется немедленное внимание и дополнительный тщательный анализ потенциальных долгосрочных последствий для данных областей.

6.1 Неисследованные этические проблемы

Все более сложные инструменты генеративного ИИ вызовут дополнительные этические проблемы, которые необходимо изучить подробно. В дополнение к разделам 2 и 3 требуется более глубокое и перспективное исследование для выявления и решения неизведанных этических проблем, освещая, по меньшей мере, следующие пять аспектов:

- **Доступ и равенство:** Системы генеративного ИИ в сфере образования могут содействовать углублению существующих различий в доступе к технологиям и образовательным ресурсам, что может дополнительно увеличивать уровень неравенства.
- **Взаимодействие между людьми:** системы генеративного ИИ в образовании могут снизить межличностное взаимодействие и критически важные социально-эмоциональные аспекты обучения.
- **Интеллектуальное развитие человека:** системы генеративного ИИ в образовании могут ограничивать автономию и свободу действий учащихся, предоставляя заранее определенные решения или уменьшая диапазон возможного обучающего опыта. Необходимо провести исследование их долгосрочного воздействия на интеллектуальное развитие учащихся.

- **Психологическое воздействие:** системы генеративного ИИ, имитирующие взаимодействие с людьми, могут оказывать неизвестное психологическое воздействие на учащихся, вызывая беспокойство относительно их когнитивного развития и эмоционального благополучия, а также относительно возможности использования этих систем для манипуляции.
- **Скрытая предвзятость и дискриминация:** поскольку в образовании разрабатываются и внедряются более сложные системы генеративного ИИ, они, вероятно, будут порождать новые предвзятости и формы дискриминации на основе данных обучения и методов, используемых в моделях. Это может привести к неизвестным и потенциально негативным результатам.

6.2 Авторское право и интеллектуальная собственность

Появление генеративного ИИ стремительно изменяет процессы создания, распространения и потребления научных, художественных и литературных произведений. Несанкционированное копирование, распространение или использование произведений, защищенных авторским правом, без разрешения правообладателя нарушает его исключительные права и может привести к юридическим последствиям. В частности, в одном из последних случаев песня, созданная с использованием генеративного ИИ и исполненная «Drake» и «The Weeknd» (Абель Тесфай), достигла миллионов слушателей, прежде чем была удалена из-за спора об авторских правах (Коскарелли, 2023). В условиях развивающейся нормативной базы требование от поставщиков генеративного ИИ признавать и защищать интеллектуальную собственность владельцев контента становится сложной задачей, и установление права собственности и оригинальности созданных

произведений представляется все более сложным. Это отсутствие прозрачности вызывает беспокойство не только в контексте защиты прав создателей и обеспечения справедливой компенсации за их интеллектуальный вклад, но также создает проблемы в образовательном контексте в отношении ответственного использования контента, созданного инструментами генеративного ИИ. Такие сложности и непрозрачность могут оказать глубокое воздействие на исследовательскую систему.

6.3 Источники контента и обучения

Инструменты генеративного ИИ изменяют способы создания и предоставления образовательного контента. В будущем контент, создаваемый в результате взаимодействия человека и ИИ, может стать основным источником знаний. Это, вероятно, еще более подорвет традиционное прямое взаимодействие учащихся с образовательным контентом, основанным на ресурсах, учебниках и учебных программах, разработанных и подтвержденных экспертами-людьми. Авторитетный внешний вид текста, созданного генеративным ИИ, может ввести в заблуждение учащихся, у которых нет достаточных предварительных знаний для распознавания неточностей или эффективного их осмысления. Вопрос о том, следует ли рассматривать взаимодействие учащихся с непроверенным контентом как «обучение», также является спорным.

Сосредоточение на агрегированной информации из вторичных источников может ограничить возможности учащихся формировать знания с применением утвержденных методов, таких как непосредственное восприятие и практический опыт, обучение путем проб и ошибок, проведение эмпирических экспериментов и развитие критического мышления. Этот подход также может нарушить социальное формирование знаний и укрепление социальных ценностей через совместные учебные практики в учебной среде.

6.4 Однородные ответы против разнообразных и творческих результатов

Генеративный ИИ ограничивает разнообразие мнений, поскольку получаемые результаты часто отражают и усиливают доминирующие точки зрения. Это приводит к возникающей гомогенизации знаний, что сдерживает плюралистическое и творческое мышление. Зависимость учителей и учащихся от инструментов генеративного ИИ для формулировки предложений может привести к стандартизации и унификации ответов, что ослабляет ценность независимого мышления и самостоятельного исследования. Потенциальная гомогенизация выражения в письменных произведениях и произведениях искусства может ограничивать воображение, креативность и альтернативные точки зрения учащихся.

Поставщики генеративного искусственного интеллекта и учителя, использующие его, должны учитывать, насколько EdGPT может быть разработан и применен для поощрения творчества, сотрудничества, критического мышления и других высших когнитивных навыков.

6.5 Переосмысление оценивания и результатов обучения

Воздействие применения генеративного ИИ для оценки выходит за рамки непосредственных опасений относительно списывания учащихся при выполнении письменных заданий. Следует признать, что генеративный ИИ способен формировать относительно структурированные статьи, эссе и впечатляющие произведения искусства. Он также может успешно справляться с некоторыми экзаменами, основанными на знаниях, в определенных предметных областях. В этой связи требуется пересмотреть содержание обучения, его цели, а также методы оценки и подтверждения уровня обучения.

Критический анализ, проводимый преподавателями, политиками, учащимися и другими заинтересованными сторонами,

должен учитывать следующие четыре категории результатов обучения:

Ценности: Ценности, необходимые для обеспечения ориентации на человека в проектировании и использовании технологий, играют ключевую роль в переосмыслении результатов обучения и их оценки в цифровую эпоху. При пересмотре целей образования важно четко определить ценности, которые руководят взаимодействием технологий с образовательным процессом. Эти ценности должны служить нормативной основой для итеративного обновления результатов обучения, их оценки и проверки, чтобы эффективно реагировать на все более широкое внедрение технологий, включая ИИ, в образовательной среде.

Базовые знания и навыки. В областях, где инструменты генеративного ИИ могут превосходить человека, учащимся всё равно необходимо обладать глубокими базовыми знаниями и навыками, такими как грамотность, умение работать с числами и математической информацией, а также основные научные знания. Необходимо регулярно пересматривать объем и характер этих фундаментальных навыков, чтобы адаптировать их ко всё более разнообразной среде, обогащенной искусственным интеллектом, в которой мы существуем.

Навыки мышления более высокого порядка. Результаты обучения должны включать навыки, необходимые для поддержки мышления высокого порядка и решения проблем в контексте сотрудничества между человеком и ИИ, а также для использования результатов, предоставляемых генеративным ИИ. Эти навыки могут включать в себя понимание роли как фактических, так и концептуальных знаний в обосновании мышления высокого порядка, а также критическую оценку контента, созданного ИИ.

Профессиональные навыки, необходимые для работы с ИИ. В сферах, где ИИ может превосходить человеческие способности и автоматизировать задачи, учащимся необходимо развивать новые

навыки, позволяющие им проектировать, использовать и взаимодействовать с инструментами генеративного ИИ. Пересмотр результатов обучения и оценки образования должны отражать профессиональные компетенции, необходимые для новых рабочих мест, созданных в результате развития ИИ.

6.6 Мыслительные процессы

Ключевой аспект долгосрочных последствий использования генеративного ИИ в образовании и исследованиях связан с эволюцией взаимодействия между человеком и машинами. Одним из центральных вопросов является возможность людей уступить базовые функции мышления и процессы усвоения навыков ИИ, сосредоточившись вместо этого на высших когнитивных процессах, опираясь на результаты, предоставленные ИИ.

Процесс написания, например, часто связан с организацией мышления. Благодаря генеративному ИИ люди теперь могут начать с уже структурированного плана, предоставленного этим инструментом, вместо того чтобы создавать его с нуля, определяя цели, объем и структуру набора идей. Некоторые эксперты описывают этот подход к использованию генеративного ИИ в текстовом творчестве как «создание текста без осознанного участия» (Чайка, 2023). По мере распространения этих новых методов, основанных на генеративном ИИ, необходимо будет адаптировать установившиеся методы обучения и оценки навыков письма. Один из возможных сценариев будущего предполагает, что обучение письму будет сконцентрировано на развитии навыков планирования и оценки, критического анализа результатов, предоставленных генеративным ИИ, а также на развитии мышления более высокого порядка. Это также может включать совместное написание на основе планов, созданных генеративным ИИ.

Заключительные комментарии

В рамках подхода, ориентированного на человека, развитие и использование ИИ должны способствовать расширению и усилению интеллектуальных и социальных способностей человека, избегая конфликтов и узурпации. Интеграция инструментов ИИ в повседневную практику поможет создать более инклюзивное и устойчивое будущее.

Для того, чтобы ИИ стал надежным партнером в образовании и исследованиях, необходимо разрабатывать и внедрять подходы, соответствующие этическим принципам, выдвинутым в «Рекомендациях ЮНЕСКО об этических аспектах ИИ» (2021 г.). Такой подход

обеспечит доверие к генеративному ИИ среди исследователей, преподавателей и учащихся.

Важно осознавать, что использование генеративного ИИ в образовании и исследованиях может вызвать изменения в существующих системах и их основах. Любые трансформации, инициированные генеративным ИИ, должны подвергаться внимательному анализу и управлению с использованием подхода, ориентированного на человека. Исключительно таким образом можно обеспечить, что потенциал ИИ будет максимально реализован в целях расширения человеческих возможностей и формирования инклюзивного цифрового будущего для всех.

Список литературы

Андерс, В. А. 2023. *Является ли использование ChatGPT мошенничеством или плагиатом, либо ни тем, ни другим, или дальновидным подходом?* Кембридж, Cell Press. Доступно по адресу: <https://doi.org/10.1016/j.patter.2023.100694> (по состоянию на 23 июня 2023 г.).

Басс Д. и Мец Р. 2023. *Сэм Альтман из OpenAI призывает Конгресс регулировать новые мощные технологии.* Нью-Йорк, Блумберг. Доступно по адресу: <https://www.bloomberg.com/news/newsletters/2023-05-17/openai-s-sam-altman-urges-congress-to-regulate-powerful-new-ai-technology> (по состоянию на 23 июня)

Бендер Э.М., Гебру Т., Макмиллан-Мейджор А. и Шмитчелл С. 2021. *Об опасностях стохастических попугаев: могут ли языковые модели быть слишком большими? Материалы конференции Ассоциации вычислительной техники (ACM) 2021 года по вопросам справедливости, подотчетности и прозрачности.* Нью-Йорк, Ассоциация вычислительной техники. Доступно по адресу: <https://doi.org/10.1145/3442188.3445922> (по состоянию на 23 июня 2023 г.).

Бове, Т. 2023. *Крупные технологические компании представляют амбициозные планы в области искусственного интеллекта в ходе финансовых отчетов, в то время как ChatGPT вызывает потрясения в отрасли. «В ближайшие несколько месяцев вы увидите много инноваций от нас».* Нью-Йорк, Фортуна. Доступно по адресу: <https://fortune.com/2023/02/03/google-meta-apple-ai-promises-chatgpt-earnings> (по состоянию на 3 июля 2023 г.).

Боммасани Р. и др. 2021. *О возможностях и рисках фундаментальных моделей.* Стэнфорд, Стэнфордский университет. Доступно по адресу: <https://crfm.stanford.edu/report.html> (по состоянию на 23 июня 2023 г.).

Бюро авторских прав США. 2023. *Руководство по регистрации авторских прав: произведения, содержащие материал, созданный искусственным интеллектом.* Федеральный реестр, Vol. 88, № 51. Вашингтон, округ Колумбия, Бюро авторских прав США (США), Библиотека Конгресса, стр. 16190–16194. Доступно по адресу: <https://www.federalregister.gov/d/2023-05321> (по состоянию на 3 июля 2023 г.).

Грань (the Verge). 2023 а. *Соучредитель OpenAI относительно предыдущего подхода компании к открытому обмену результатами исследований: «Мы допустили ошибку».* Вашингтон, округ Колумбия, Вокс Медиа (Vox Media). Доступно по адресу: <https://www.theverge.com/2023/3/15/23640180/openai-gpt-4-launch-closed-research-ilya-sutskever-interview> (по состоянию на 1 августа 2023 г.).

---- . 2023 б. *Генеральный директор OpenAI Сэм Альтман о GPT-4: «Люди предвкушают разочарование, поэтому в итоге они его и получают».* Вашингтон, округ Колумбия, Вокс Медиа (Vox Media). Доступно по адресу: <https://www.theverge.com/23560328/openai-gpt-4-rumor-release-date-sam-altman-interview> (по состоянию на 1 августа 2023 г.).

Гугл. 2023 а. *Рекомендации по регулированию ИИ.* Маунтин-Вью, Гугл (Google). Доступно по адресу: <https://ai.google/static/documents/recommendations-for-regulation-ai.pdf> (по состоянию на 23 июня 2023 г.).

---- . 2023 б. *Технический отчет о PaLM 2.* Маунтин-Вью, Гугл (Google). Доступно по адресу: <https://doi.org/10.48550/arXiv.2305.10403> (по состоянию на 20 июля 2023 г.).

Двиведи, Ю.К., Кшетри, Н., Хьюз, Л., Слэйд, Э.Л., Джеярадх, А., Кар, АК, Баабдулла, А.М., Куханг, А, Рагхаван, В., Ахуджа, М., Албанна, Х., Альбашрави, М.А., Ал - Бусаиди А.С., Балакришнан Дж., Барлетт Ю., Басу С., Бозе И., Брукс Л., Бухалис Д., Картер Л., Чоудхури С. и др., Крик, Т., Каннингем, С.В., Дэвис, Г.Х., Дэвисон, Р.М., Де, Р., Деннехи, Д., Дуан, Ю., Дубей, Р., Двиведи, Р., Эдвардс, Дж.С. Флавиан, К., Голд, Р., Гровер В., Ху, М.-К., Янссен, М., Джонс, П., Джунглас, И., Корана, С., Краус, С., Ларсен, КР, Латрейль, П., Лаумер, С., Малик, Ф.Т., Мардани, А., Мариани, М., Митас, С., Могаджи, Э., Норт Хорн, Дж., О'Коннор, С., Окумус, Ф., Пагани, М., Пандей, Н., Папаяннидис, С., Паппас, И.О., Патхак, Н., Прис-Хедже, Ж., Раман, Р., Рана, Н.П., Рем, С.-В., Рибейро-Наваррете, С., Рихтер, А., Роу, Ф., Саркер, С., Шталь, БК, Тивари, М.К., ван дер Аалст, В., Венкатеш, В., Вилья, Г., Уэйд, М., Уолтон, П., Вирц Дж. и Райт Р. 2023. *Междисциплинарный анализ возможностей, проблем и последствий генеративного диалогового искусственного интеллекта для исследований, практики и политики*. Международный журнал информационного менеджмента, Vol. 71. Амстердам, Эльзевир, стр. 102642. Доступно по адресу: <https://doi.org/10.1016/j.ijinfomgt.2023.102642> (по состоянию на 25 августа 2023 г.).

Джаннини, С. 2023. *Генеративный искусственный интеллект и будущее образования*. Париж, ЮНЕСКО. Доступно по адресу: <https://unesdoc.unesco.org/ark:/48223/pf0000385877> (по состоянию на 29 августа 2023 г.).

Е2Аналитик. 2023. *GPT-4: все, что вы хотели знать о новой модели искусственного интеллекта OpenAI*. Сан-Франциско, Медиум. Доступно по адресу: <https://medium.com/predict/gpt-4-everything-you-want-to-know-about-openais-new-ai-model-a5977b42e495> (по состоянию на 1 августа 2023 г.).

Европейская комиссия. 2021. *Установление гармонизированных правил в области искусственного интеллекта (Закон об искусственном интеллекте) и внесение изменений в некоторые законодательные акты Европейского союза*. Брюссель, Европейская комиссия. Доступно по адресу: <https://artificialintelligenceact.eu> (по состоянию на 23 июня 2023 г.).

Евросоюз. 2016. *Регламент (ЕС) 2016/679 Европейского парламента и Совета от 27 апреля 2016 года о защите физических лиц в связи с обработкой персональных данных и о свободном передвижении таких данных, а также об отмене Директивы 95/46/ЕС (Общий регламент по защите данных)*. Брюссель, Официальный журнал Европейского Союза. Доступно по адресу: <http://data.europa.eu/eli/reg/2016/679/oj> (по состоянию на 23 июня 2023 г.).

Коскарелли, Дж. 2023. *Искусственный интеллект с фейковыми «Drake» и «The Weeknd» потрясает музыкальный мир*. Нью-Йорк, Нью-Йорк Таймс. Доступно по адресу: <https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html> (по состоянию на 30 августа 2023 г.).

Лин, Б. 2023. *ИИ создает угрозы безопасности быстрее, чем компании могут справиться с этим вызовом*. Нью-Йорк, Уолл-стрит Джорнал (The Wall Street Journal). Доступно по адресу: <https://www.wsj.com/articles/ai-is-generating-security-risks-faster-than-companies-can-keep-up-a2bdedd4> (по состоянию на 25 августа 2023 г.).

Марвала, Т. 2023. *Смещение алгоритма — использование синтетических данных следует рассматривать как последнюю меру при обучении систем искусственного интеллекта*. Университет ООН. Доступно по адресу: <https://unu.edu/article/algorithm-bias-synthetic-data-should-be-option-last-resort-when-training-ai-systems> (по состоянию на 31 июля 2023 г.).

Маркус, Г. 2022. *В надежде на лучшее в ходе эволюции искусственного интеллекта*. Материалы Ассоциации вычислительной техники (ACM), Vol. 66, № 4. Нью-Йорк, Ассоциация вычислительной техники (ACM). Доступно по адресу: <https://doi.org/10.1145/3583078> (по состоянию на 23 июня 2023 г.).

Мерфи Келли, С. 2023. *Microsoft внедряет технологию ChatGPT в Word, Excel и Outlook*. Атланта, CNN. Доступно по адресу: <https://edition.cnn.com/2023/03/16/tech/openai-gpt-microsoft-365/index.html> (по состоянию на 25 августа 2023 г.).

Мец, К. 2021. *Кто осуществляет мониторинг за тем, чтобы системы ИИ не демонстрировали признаки расовой предвзятости?* Нью-Йорк, Нью-Йорк Таймс. Доступно по адресу: <https://www.nytimes.com/2021/03/15/technology/artificial-intelligence-google-bias.html> (по состоянию на 23 июня 2023 г.).

Назарецкий Т., Ариели М., Чукурова М. и Александрон Г. 2022 б. Доверие учителей к образовательным технологиям, основанным на искусственном интеллекте, и программа профессионального развития для его усовершенствования. *Британский журнал образовательных технологий*, Vol. 53, № 4. Хобокен, Нью-Джерси, Уайли, стр. 914–931. Доступно по адресу: <https://doi.org/10.1111/bjet.13232> (по состоянию на 1 августа 2023 г.).

Назарецкий Т., Чукурова М. и Александрон Г. 2022 а. Инструмент для измерения доверия учителей образовательным технологиям на основе искусственного интеллекта. *12-я Международная конференция по аналитике обучения и знаниям*. Ванкувер, Ассоциация вычислительной техники, стр. 55–66.

Окампо, Ю. 2023. *Сингапур запускает правительственный облачный кластер искусственного интеллекта*. Сингапур, OpenGov Asia. Доступно по адресу: <https://opengovasia.com/singapore-unveils-ai-government-cloud-cluster> (по состоянию на 25 августа 2023 г.).

Open AI. 2018. *ИИ и вычислительные мощности*. Сан-Франциско, OpenAI. Доступно по адресу: <https://openai.com/research/ai-and-compute> (по состоянию на 23 июня 2023 г.).

----. 2023. *Аспекты, которые следует учитывать педагогам при использовании ChatGPT*. Сан-Франциско, OpenAI. Доступно по адресу: <https://platform.openai.com/docs/chatgpt-education> (по состоянию на 23 июня 2023 г.).

Попли, Н. 2023. *Работа в сфере искусственного интеллекта с зарплатой до 335 тысяч долларов, не требующая образования в области компьютерной инженерии*. Нью-Йорк, Тайм США (TIME USA). Доступно по адресу: <https://time.com/6272103/ai-prompt-engineer-job> (по состоянию на 23 июня 2023 г.).

Рассел Групп (Russell Group), 2023. *Принципы Russell Group по использованию инструментов генеративного искусственного интеллекта в образовании*. Кембридж, Рассел Групп. Доступно по адресу: https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf (по состоянию на 25 августа 2023 г.).

Руз, К. 2022. *Изображение, сгенерированное искусственным интеллектом, было удостоено художественной премии. Художники выразили свое недовольство*. Нью-Йорк, Нью-Йорк Таймс. Доступно по адресу: <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html> (по состоянию на 23 июня 2023 г.).

Стэнфордский Университет. 2019. *Отчет об индексе искусственного интеллекта*. Стэнфорд, Стэнфордский университет. Доступно по адресу: <https://hai.stanford.edu/ai-index-2019> (по состоянию на 23 июня 2023 г.).

----. 2023. *Отчет об индексе искусственного интеллекта*. Стэнфорд, Стэнфордский университет. Доступно по адресу: <https://hai.stanford.edu/research/ai-index-2023> (по состоянию на 23 июня 2023 г.).

Тлили А., Шехата Б., Агиеманг Адарква М., Бозкурт А., Хики Д.Т., Хуанг Р. и Агиеманг Б. *Что, если дьявол - мой ангел-хранитель: ChatGPT как кейс-исследование использования чат-ботов в образовании*. Умная среда обучения, Vol. 10, № 15. Берлин, Шпрингер. Доступно по адресу: <https://doi.org/10.1186/s40561-023-00237-x> (по состоянию на 23 июня 2023 г.).

Управление киберпространством Китая, 2023 г. *Публичный запрос мнений по «Мерам по управлению услугами генеративного искусственного интеллекта (проект для комментариев)»*. Управление киберпространством Китая (CAC), Пекин (на китайском языке). Доступно по адресу: http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm (по состоянию на 19 июля 2023 г.).

----- 2023 в. Временные меры по управлению сервисами генеративного искусственного интеллекта. Управление киберпространством Китая (САС), Пекин (на китайском языке). Доступно по адресу: http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm (по состоянию на 19 июля 2023 г.)

Федеральная торговая комиссия. 1998. *Закон о защите конфиденциальности детей в Интернете 1998 года*. Вашингтон, округ Колумбия, Федеральная торговая комиссия. Доступно по адресу: <https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa> (по состоянию на 4 сентября 2023 г.).

Чайка, К. 2023. *Мой отчет по написанию с использованием искусственного интеллекта*. Нью-Йорк, Нью-Йоркер. Доступно по адресу: <https://www.newyorker.com/cultural/infinite-scroll/my-ai-writing-robot> (по состоянию на 1 августа 2023 г.).

Чен Л., Захария М. и Цзоу Дж. 2023. *Как поведение ChatGPT меняется с течением времени?* Итака, arXiv. Доступно по адресу: <https://arxiv.org/pdf/2307.09009> (по состоянию на 31 июля 2023 г.).

ЮНЕСКО. 2019. *Пекинский консенсус по искусственному интеллекту и образованию*. Париж, ЮНЕСКО. Доступно по адресу: <https://unesdoc.unesco.org/ark:/48223/pf0000368303> (по состоянию на 3 июля 2023 г.).

---- . 2022 а. *Рекомендации по этике искусственного интеллекта*. Париж, ЮНЕСКО. Доступно по адресу: <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (по состоянию на 3 июля 2023 г.).

---- . 2022 б. *Технологии искусственного интеллекта в образовании: перспективы и последствия*. Париж, ЮНЕСКО. Доступно по адресу: <https://unesdoc.unesco.org/ark:/48223/pf0000376709> (по состоянию на 23 июня 2023 г.).

---- . 2022 в. *Учебные программы по ИИ для общего образования. Картирование одобренных правительством учебных программ по ИИ*. Париж, ЮНЕСКО. Доступно по адресу: <https://unesdoc.unesco.org/ark:/48223/pf0000380602> (по состоянию на 20 июля 2023 г.).

---- . 2022 г. *Руководящие принципы для разработки политики и генеральных планов по ИКТ в образовании*. Париж, ЮНЕСКО. Доступно по адресу: <https://unesdoc.unesco.org/ark:/48223/pf0000380926> (по состоянию на 31 июля 2023 г.).

---- . 2023 а. *Искусственный интеллект: ЮНЕСКО призывает все правительства незамедлительно внедрить Глобальные этические рамки*. Париж, ЮНЕСКО. Доступно по адресу: <https://www.unesco.org/en/articles/artificial-intelligence-unesco-calls-all-governments-implement-global-ethical-framework-without> (по состоянию на 3 июля 2023 г.).

---- . 2023 б. *Картирование и анализ правительственных стратегий по регулированию и содействию творческому использованию генеративного ИИ*. Неопубликовано.

---- . 2023 в. *Исследование использования ИИ правительством в качестве общественного блага в сфере образования*. Неопубликовано (представлено в ЮНЕСКО).

---- . 2023. *Всемирный доклад по мониторингу образования, 2023 год: технологии в образовании — инструмент на чьих условиях?* Париж, Рабочая группа по Всемирному докладу по мониторингу образования. Доступно по адресу: <https://unesdoc.unesco.org/ark:/48223/pf0000385723> (по состоянию на 25 августа 2023 г.).

---- . 2023. *ChatGPT и искусственный интеллект в высшем образовании. Краткое руководство*. Каракас, Международный институт ЮНЕСКО по высшему образованию в Латинской Америке и Карибском бассейне. Доступно по адресу: <https://unesdoc.unesco.org/ark:/48223/pf0000385146> (по состоянию на 25 августа 2023 г.).

СНОСКИ

- 1 Модели генеративного ИИ стали доступны исследователям и другим заинтересованным лицам гораздо раньше, чем ChatGPT. Например, в 2015 году Google запустил «DeepDream» (<https://en.wikipedia.org/wiki/DeepDream>).
- 2 См. <https://chat.openai.com>
- 3 Объяснение методов и технологий искусственного интеллекта и их взаимосвязи см. в ЮНЕСКО, 2022*b*, стр. 8–10.
- 4 Обратите внимание: поскольку генеративный искусственный интеллект все еще относительно новый, разные компании часто используют эти термины по-разному и иногда применяют различные терминологии для обозначения одних и тех же концепций.
- 5 Существует опасение, что данные, используемые для обучения будущих версий OpenAI GPT, могут содержать значительные объемы текста, созданного предыдущими версиями GPT. Этот цикл обращения к себе может исказить данные обучения и, таким образом, создать риск для потенциальных возможностей будущих моделей GPT.
- 6 Компания OpenAI, разработавшая GPT, не предоставила подробной публичной информации о GPT-4 в данной таблице (The Verge, 2023*a*). Фактически, количество параметров было опровергнуто генеральным директором OpenAI (The Verge, 2023*b*). Тем не менее эти цифры были опубликованы в некоторых СМИ (например, см. E2Analyst, 2023). В любом случае, основной вывод заключается в том, что GPT-4 построен на значительно более обширном корпусе данных и использует намного больше параметров по сравнению с GPT-3.
- 7 См. <https://crfm.stanford.edu/2023/03/13/alpaca.html>.
- 8 См. <https://bard.google.com>.
- 9 См. <https://writersonic.com/chat>.
- 10 См. <https://yiyan.baidu.com/welcome>.
- 11 См. <https://huggingface.co/chat>.
- 12 См. <https://www.jasper.ai>.
- 13 См. <https://ai.facebook.com/blog/large-language-model-llama-meta-ai>.
- 14 См. <https://open-assistant.io>.
- 15 См. <https://www.alizila.com/alibaba-cloud-debuts-generative-ai-model-for-corporate-users>.
- 16 См. <https://you.com>.
- 17 См. <https://www.chatpdf.com>.
- 18 См. <https://elicit.org>.
- 19 См. <https://www.perplexity.ai>.
- 20 См. <https://tools.zmo.ai/webchatgpt>.
- 21 См. <https://www.compose.ai>.
- 22 См. <https://www.teamsmart.ai>.
- 23 См. <https://wiseone.io>.
- 24 См. <https://www.microsoft.com/en-us/bing>.
- 25 См. <https://www.craiyon.com>.
- 26 См. <https://openai.com/product/dall-e-2>.
- 27 См. <https://dream.ai/create>.
- 28 См. <https://www.fotor.com/features/ai-image-generator>.
- 29 См. <https://www.midjourney.com>.
- 30 См. <https://creator.nightcafe.studio>.
- 31 См. <https://writersonic.com/photosonic-ai-art-generator>.
- 32 См. <https://elai.io>.

- 33 См. <https://www.gliacloud.com>.
- 34 См. <https://pictory.ai>.
- 35 См. <https://runwayml.com>.
- 36 См. <https://www.aiva.ai>.
- 37 См. <https://boomy.com>.
- 38 См. <https://sounddraw.io>.
- 39 См. <https://www.voicemod.net/text-to-song>.
- 40 См. <https://openai.com/research/gpt-4>.
- 41 См. <https://www.educhat.top> и <https://www.mathgpt.com>.
- 42 См. <https://www.educhat.top>.
- 43 См. <https://www.mathgpt.com>.
- 44 Есть несколько исключений, например Hugging Face, группа, занимающаяся разработкой искусственного интеллекта с открытым исходным кодом.
- 45 См., например, Google (2023a) и OpenAI (Bass and Metz, 2023).
- 46 Об одном проекте по регулированию ИИ см. в проекте Закона Европейской комиссии об ИИ (2021 г.).
- 47 Обзор базировался на данных, собранных в рамках опроса, проведенного ЮНЕСКО среди 193 государств-членов организации относительно использования искусственного интеллекта в сфере образования правительствами. (ЮНЕСКО, 2023 г.), Обсерватории ОЭСР по вопросам политики в сфере ИИ, на основе Отчета об индексе искусственного интеллекта Стэнфордского университета (Стэнфордский университет, 2023), а также с помощью первичной информации, полученной от группы международных экспертов.
- 48 См. <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>.
- 49 Из обзора, по состоянию на апрель 2023 года, следующие страны опубликовали национальные стратегии в области искусственного интеллекта: Аргентина, Австралия, Австрия, Бельгия, Бенин, Бразилия, Канада, Болгария, Чили, Китай, Колумбия, Кипр, Чехия, Дания, Египет, Эстония, Финляндия, Франция, Германия, Венгрия, Исландия, Индия, Индонезия, Ирландия, Италия, Япония, Иордания, Латвия, Литва, Люксембург, Малайзия, Мальта, Маврикий, Мексика, Нидерланды, Норвегия, Новая Зеландия, Оман, Перу, Польша, Португалия, Филиппины, Катар, Республика Корея, Румыния, Российская Федерация, Саудовская Аравия, Сербия, Сингапур, Словения, Испания, Швеция, Таиланд, Турция, Тунис, Объединенные Арабские Эмираты, Великобритания, США, Уругвай и Вьетнам. Кроме того, некоторые страны включили стратегии ИИ в более широкие стратегии в области ИКТ или цифровых технологий, в том числе Алжир, Ботсвана, Казахстан, Кения, Сьерра-Леоне, Словакия, Швейцария и Уганда.
- 50 Согласно обзору всех национальных стратегий в области ИИ (ЮНЕСКО, 2023b), более 40 стратегий включают специальные разделы, посвященные вопросам этики.
- 51 Согласно обзору всех национальных стратегий ИИ (ЮНЕСКО, 2023b), около 45 стратегий содержат разделы, посвященные вопросам образования.
- 52 См. <https://openai.com/policies/terms-of-use>.
- 53 В определенных странах преподаватели имеют ассистентов (Teaching Assistants), чья функция заключается в ответах на вопросы отдельных учащихся, касающихся учебного материала. Генеративный искусственный интеллект может быть задействован для создания генеративного аналога ассистента, который обеспечивает поддержку учащимся и преподавателям, но при этом могут возникнуть негативные аспекты, такие как вопросы, связанные с социальным взаимодействием в классе.



unesco

Организация
Объединенных Наций по
вопросам образования,
науки и культуры

Руководство по использованию генеративного искусственного интеллекта в образовании и научных исследованиях

Данное руководство направлено на поддержку разработки соответствующих правил, политик и программ развития человеческого потенциала с целью обеспечения того, чтобы генеративный искусственный интеллект (ИИ) стал инструментом, действительно приносящим пользу и расширяющим возможности учителей, учащихся и исследователей. В нем объясняются методы искусственного интеллекта, используемые в генеративном ИИ, а также приводится список моделей GPT, которые становятся общедоступными, особенно модели с открытым исходным кодом. Руководство также открывает обсуждение о появлении EdGPT – моделей генеративного ИИ, обученных на конкретных данных для образовательных целей. Кроме того, в нем суммируются некоторые ключевые противоречия вокруг генеративного ИИ – от ухудшения цифрового неравенства до гомогенизации мнений, от более сложных дипфейков до вопросов авторского права. Исходя из гуманистического подхода, руководство предлагает ключевые шаги по регулированию инструментов генеративного ИИ, включая обязательную защиту конфиденциальности данных и установку возрастных ограничений для самостоятельного взаимодействия с платформами генеративного ИИ. Для обеспечения правильного использования инструментов в сфере образования и исследований, в руководстве затронуты вопросы этики и проектирования педагогического процесса и заложен подход, ориентированный на человека и учитывающий возрастные особенности.

