

Генеративный ИИ в европейской науке: инструкция по применению

Технологии генеративного искусственного интеллекта (GenAI)¹ призваны повысить производительность исследовательского процесса и способствовать достижению мирового лидерства ЕС в исследованиях и разработках. Усилия европейских законодателей также направлены на управление рисками, связанными с использованием GenAI. Институт статистических исследований и экономики знаний (ИСИЭЗ) НИУ ВШЭ изучил европейские рекомендации по регулированию применения ИИ для целей науки.

Мировой контекст регулирования сферы ИИ

Вопросы регулирования искусственного интеллекта (ИИ) давно включены в повестку многих стран и международных организаций. В 2016 г. в США [приняли](#) первый законодательный акт в этой сфере, а к 2022 г. их [количество](#) в мире возросло до 37. В 2023 г. страны G7 [согласовали](#) международный кодекс и руководящие принципы для компаний-разработчиков ИИ, также на уровне Евросоюза и 28 других государств была [подписана](#) Декларация Блетчли по безопасному применению ИИ ([подробнее](#) – в [материале](#) «[Регулирование искусственного интеллекта: первые шаги](#)»).

В 2024 г. международные законодатели сфокусировались на обеспечении ответственного и этичного использования ИИ, в том числе в науке. Так, ОЭСР [обновила](#) рекомендацию по ИИ (первый межправительственный стандарт по ответственному управлению надежным ИИ), разработала [подход](#) к классификации систем ИИ и [запустила](#) проект по мониторингу применения кодекса G7 для компаний-разработчиков ИИ. ЮНЕСКО приняла два документа: [рекомендацию](#) по этике ИИ и [руководство](#) по использованию GenAI в образовании и научных исследованиях. Всемирная организация здравоохранения (ВОЗ) [обновила](#) руководство «Этика и управление ИИ в здравоохранении», а Совет Европы [разработал](#) Рамочную конвенцию об ИИ, правах человека, демократии и верховенстве права.

Ускоренное продвижение ИИ в европейскую науку

В январе 2024 г. Европейская комиссия (ЕК) в рамках создания европейского центра экспертизы в области ИИ и реализации положений Закона об ИИ ([AI Act](#)) учредила надзорный орган регулирования ИИ ([AI Office](#))². Он осуществляет деятельность на трех уровнях: политическом, нормативно-правовом и контроля безопасности, а также отвечает за координацию европейской политики в сфере ИИ и взаимодействие как со странами – членами ЕС, так и с более широким экспертным и научным сообществом.

Европейская группа главных научных советников (European Group of Chief Scientific Advisors) предложила учредить своего рода «ЦЕРН для ИИ» (CERN for AI) – Европейский распределенный институт по использованию ИИ в науке (European Distributed Institute for AI in Science, EDIRAS), призванный обеспечить исследователям, использующим ИИ, особые условия (доступ к высокопроизводительным вычислительным мощностям, устойчивой облачной инфраструктуре, высококачественным данным и специализированным программам обучения). На его создание и деятельность в течение ближайших пяти-семи лет предполагается [выделить](#) до 100 млрд евро (104.8 трлн руб.)³ из бюджета формируемого Европейского научного совета по ИИ в науке.

В апреле 2024 г. ЕК выпустил [рекомендации](#) по поддержке применения ИИ в исследованиях, в том числе касающиеся повышения доступности ресурсов для проведения научных исследований с использованием ИИ, улучшения условий их финансирования, развития кадрового потенциала в данной сфере, а также повышения качества получаемых результатов (рис. 1). Документ учитывает результаты оценки европейской и международной политики в области ИИ, анализ возможностей и барьеров, препятствующих масштабированию использования ИИ в науке.

¹ Разновидность искусственного интеллекта, который может создавать новый контент.

² Аналогичные органы были созданы в США ([AIRC](#), [AISI](#), [CISA](#)), Великобритании ([NCSC](#), [AISI](#)) и Австралии ([ACSC](#)).

³ Здесь и далее суммы в рублях представлены по курсу ЦБ РФ на 23.10.2024 г., равному 104.87 за 1 евро.

Рис. 1. Рекомендации Европейской комиссии по поддержке применения ИИ в исследованиях

Рекомендация 1. Гибкое финансирование и масштабирование использования ИИ в исследованиях	Рекомендация 2. Качество и доступность	Рекомендация 3. Единая инфраструктура	Рекомендация 4. Таланты, равенство и справедливость
<ul style="list-style-type: none"> - внедрение гибких механизмов финансирования исследований с использованием ИИ; - финансирование инициатив и практик по оценке воздействия ИИ на окружающую среду; - поддержка разработки этического ИИ; - финансирование исследователей, изучающих и использующих ИИ в своей деятельности; - создание распределенной сети по использованию ИИ для проведения совместных исследований; - финансирование разработки программного, аппаратного и кадрового обеспечения для использования ИИ в науке (фундаментальных моделей и больших языковых моделей, лабораторной робототехники, цифровых научных ассистентов); - мониторинг развития ИИ и политики использования ИИ в исследованиях; - оценка потенциальных угроз и рисков применения ИИ в науке; - разработка стандартов и нормативно-правового регулирования; - предоставление информации, рекомендаций и инструментов ученым для оказания им поддержки в решении вопросов, связанных с правовыми, социальными и этическими аспектами и нормами, а также с управлением ИИ; - создание инструментов мониторинга потенциальных злоупотреблений ИИ. 	<ul style="list-style-type: none"> - обеспечение равного доступа ученых к наборам исследовательских данных; - поддержка междисциплинарных исследований; - совершенствование стандартов качества исследований с применением ИИ; - оценка ограничений использования ИИ в исследованиях; - внедрение системы водяных знаков для маркировки данных, созданных с помощью ИИ*; - госзакупка на основе ЧГП у частных компаний данных для проведения исследований с использованием ИИ. 	<ul style="list-style-type: none"> - развитие единой исследовательской инфраструктуры использования систем ИИ; - поддержка компаний, МСП, ЧГП, предоставляющих исследователям услуги ИИ; - создание на базе университетов и исследовательских организаций школ ИИ; - разработка и включение в образовательные программы вузов модулей по ИИ; - создание альянсами университетов совместной аспирантуры в сфере ИИ. 	<ul style="list-style-type: none"> - наращивание кадрового потенциала в сфере ИИ; - применение механизмов привлечения и удержания талантов (академические преимущества, инструменты для повышения качества исследований, доступ к вычислительным инфраструктурам и данным); - поддержка междисциплинарного научно-технологического сотрудничества ученых; - поддержка исследований, не приносящих существенного дохода, но служащих общественным интересам; - разработка рекомендаций, ограничивающих использование инструментов ИИ в рецензировании или оценке исследовательских грантов, для преодоления рисков фальсификации результатов исследований.

* Встраивание в генерируемые модели ИИ контент уникального сигнала с целью маркировки такого ИИ-контента.

Источник: ИСИЭЗ НИУ ВШЭ по данным Европейской комиссии.

В рамках инициативы [EuroHPC](#) по формированию европейской суперкомпьютерной экосистемы мирового класса [планируется](#) к 2025 г. создать «фабрики ИИ» (см. справку) и обеспечить к ним доступ разным категориям пользователей, включенных в инновационную экосистему ЕС, – от исследовательских институтов до промышленных предприятий и учреждений госсектора. При этом эксклюзивный доступ будет предоставляться ИИ-стартапам и МСП для тестирования и проверки моделей GenAI. Общие инвестиции на создание и развитие «фабрик ИИ» из средств программ Horizon Europe, Digital Europe и других, в том числе частных, источников на период до 2027 г. могут составить до 4 млрд евро (419.5 млрд руб.).

«Фабрики ИИ» (AI Factory) – объединенные в открытую экосистему специализированные суперкомпьютеры, используемые для разработки, тестирования, оценки и валидации генеративных моделей ИИ, требующих значительных вычислительных мощностей. Выступают центрами развития приложений ИИ в ключевых для ЕС областях (здравоохранение, энергетика, промышленность, метеорология и др.). На текущий момент закуплены девять суперкомпьютеров, три из которых входят в топ-10 самых мощных в мире. ЕС покрывает организациям, размещающим на своей базе «фабрики ИИ», до 50% стоимости суперкомпьютеров и до 50% эксплуатационных расходов. Также запланировано грантовое финансирование на разработку микропроцессоров и передовой экспериментальной суперкомпьютерной платформы, оптимизированной для работы с моделями ИИ, приобретение / модернизацию суперкомпьютеров, создание сервисов и развитие навыков.

Источник: ИСИЭЗ НИУ ВШЭ по данным Европейской комиссии.

В рамках еще одной инициативы – [GenAI4EU](#) – с бюджетом 500 млн евро (52.4 млрд руб.) до 2027 г. предполагается стимулировать использование GenAI в 14 направлениях: в области естественных наук, здравоохранения, биотехнологий, сельского хозяйства, энергетики, производства и машиностроения, изменения климата и экологической устойчивости, мобильности, кибербезопасности, разработки виртуальных миров и цифровых двойников, а также в сфере освоения космоса.

Преимущества и риски использования GenAI в науке

Использование GenAI для анализа и обработки данных значительно ускоряет исследовательский процесс в части обобщения литературы, обнаружения закономерностей и планирования экспериментов. Генеративные ИИ-модели [позволяют](#) структурировать большие массивы данных, создавать новую информацию на основе имеющейся, строить прогнозы, повышать качество получаемых результатов и снижать стоимость исследований за счет оптимизации вычислительных ресурсов.

Европейский исследовательский совет [отмечает](#), что наиболее очевидные преимущества использования GenAI заключаются в более эффективном решении рутинных или трудоемких задач низкого уровня, устранении языковых барьеров, автоматической обработке больших массивов данных и извлечении из них релевантной теме исследования информации, в общем повышении производительности и расширении форматов представления научных результатов разным аудиториям (рис. 2).

Рис. 2. Основные преимущества и риски использования GenAI в науке

Преимущества	Риски
<ul style="list-style-type: none"> • автоматизация и оптимизация поиска информации и других аспектов исследовательской деятельности; • систематизация, обработка, анализ и интерпретация больших данных; • поддержка многоязычного перевода; • моделирование и симуляция сложных систем; • упрощение, ускорение и коррекция процессов кодирования; • прогностическое управление GenAI; • разработка гипотез, идей, концепций, методов изучения, научных теорий; • снижение эксплуатационных и ремонтных расходов, сокращение времени на физическую сборку и тестирование за счет применения ИИ-технологий; • снижение уровня плагиата; • подготовка обзоров из разных источников и на основе анализа больших данных. 	<ul style="list-style-type: none"> • низкий уровень проработанности регулирования, стандартизации и оценки использования GenAI; • кибербезопасность, мошенничество, дезинформация, некорректное использование GenAI; • незрелость и неготовность систем контроля и защиты (технических и нетехнических) GenAI; • языковые ограничения; • искажение, нарушение конфиденциальности обучающих данных; • ошибки, неточности, галлюцинации ИИ-моделей, недостаточная предсказуемость GenAI; • нарушение прав интеллектуальной собственности, искажение информации об источниках и авторах; • закрытость доступа, плата за доступ; • монополизация GenAI коммерческим сектором, концентрация передовых исследований и разработок в частных компаниях; • скрытие информации об использовании GenAI; • распространение некачественных результатов исследований; • ограниченная воспроизводимость, прозрачность и интерпретируемость получаемых результатов; • увеличение объема плагиата, затруднение его идентификации; • подрыв доверия к научным исследованиям.

Источник: ИСИЭЗ НИУ ВШЭ по материалам ОЭСР, ЕК и других организаций.

Несмотря на существенные преимущества, GenAI имеет ряд технических ограничений и несет определенные риски. Отслеживающая инциденты, связанные со злоупотреблением ИИ, инициатива [AIAIC](#) фиксирует с 2012 г. 26-кратное увеличение их числа. ЕК в качестве основных рисков отмечает непрозрачность ИИ-моделей, вероятность их неэтичного, навязчивого, манипулятивного и дискриминационного использования, генерацию неточных знаний, распространение дезинформации и дипфейков. Сложности нередко возникают в отношении прав интеллектуальной собственности (ИС) на произведения, созданные с помощью GenAI. Открытость и управляемость языковых моделей в виде чатботов (например, ChatGPT от OpenAI, BlenderBot3 от Meta, Bard от Google) могут оборачиваться предвзятостью и рисками потенциального раскрытия персональной информации. Рост чрезмерной зависимости от инструментов GenAI в перспективе может привести к снижению уровня навыков исследователей, а также к росту зависимости от поставщиков ИИ-услуг. Перечисленные и другие риски, которые могут проявляться по мере ускоряющегося развития систем GenAI, вызывают у европейских политиков опасения, что сектор исследований, как один из наиболее чувствительных, может оказаться под угрозой.

Рекомендации ЕС для ученых, организаций и инвесторов по вопросам управления ИИ

Важным направлением регулирования GenAI является разработка этических принципов, охватывающих такие области, как конфиденциальность, защита данных, алгоритмическая справедливость и предотвращение неправомерного использования.

В марте 2024 г. ЕК выпустила [руководство](#) по ответственному использованию GenAI. Документ содержит рекомендации ученым, научно-исследовательским организациям и финансовым институтам по обеспечению прозрачности, конфиденциальности и соблюдению прав ИС при применении ИИ. За учеными как за авторами научных работ предлагается закрепить ответственность за генерируемый с помощью ИИ контент. Они обязаны будут уведомлять о применении инструментов GenAI, обеспечивать воспроизводимость и надежность получаемых таким образом результатов, нести ответственность за использование для генерации контента неопубликованных или конфиденциальных данных, включая персональные данные третьих лиц, задействованных без их согласия. Научно-исследовательским организациям рекомендуется переходить к суверенным центрам обработки данных, локальным или облачным инструментам GenAI под собственным управлением, также им вменяется ответственность за обеспечение кибербезопасности при работе их сотрудников с ИИ-моделями; контроль за созданием и использованием научных результатов, полученных с использованием GenAI; разработка этических и юридических требований к их применению. Финансирующие организации должны будут обеспечивать прозрачные механизмы финансирования исследований с использованием GenAI, обязывать грантополучателей информировать о фактах такого использования и соблюдать стандарты конфиденциальности.

Резюме

За последний год GenAI эволюционировал в один из важных факторов обеспечения национальной конкурентоспособности и цифрового суверенитета целого ряда ведущих стран. При этом глобальное регулирование отстает от темпов технологического прогресса.

Несмотря на существенные преимущества использования GenAI (в виде сокращения расходов, ускорения процессов разработки и внедрения научных результатов, решения комплексных задач с использованием больших массивов данных), его распространение несет в себе и угрозы, касающиеся возможного искажения научных результатов, что часто невозможно проверить (сами ученые не всегда относятся критически к генерируемому контенту и перепроверяют данные).

Особые опасения вызывает непрозрачность применения GenAI: непредставление подробных сведений об использовании данного инструмента, его влиянии на процесс исследования, проблемы с интерпретируемостью и воспроизводимостью результатов. Неравенство в доступе к ресурсам обуславливает превосходство технологических гигантов в создании моделей GenAI над академическим сообществом, которое ограничено в возможностях использования больших наборов данных и вычислительных мощностей.

Для развития исследований с применением ИИ и одновременно обеспечения единых стандартов безопасности, контроля рисков и управления инструментами GenAI Евросоюз формирует систему соответствующих надзорных органов, институтов и программ господдержки.



Источники: официальные доклады и веб-сайты ОЭСР, Европейской комиссии, Европейского союза; результаты проекта «Комплексное научно-методологическое и информационно-аналитическое сопровождение разработки и реализации государственной научной, научно-технической политики» тематического плана научно-исследовательских работ, предусмотренных Государственным заданием НИУ ВШЭ.

■ Материал подготовили **Елена Сабельникова, Михаил Гершман**

Данный материал НИУ ВШЭ может быть воспроизведен (скопирован) или распространен в полном объеме только при получении предварительного согласия со стороны НИУ ВШЭ (обращаться issek@hse.ru). Допускается использование частей (фрагментов) материала при указании источника и активной ссылки на интернет-сайт ИСИЭЗ НИУ ВШЭ (issek.hse.ru), а также на авторов материала. Использование материала за пределами допустимых способов и/или указанных условий приведет к нарушению авторских прав.