ЭКСПРЕСС-ИНФОРМАЦИЯ

НАУКА
ТЕХНОЛОГИИ
ИННОВАЦИИ

ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Институт статистических исследований
и экономики знаний

Дата выпуска
12.01.2024

Регулирование искусственного интеллекта: первые шаги

Стремительный рост популярности генеративных языковых моделей (в частности, ChatGPT) в мире актуализировал дискуссии о необходимости регулировать искусственный интеллект (ИИ). Институт статистических исследований и экономики знаний (ИСИЭЗ) НИУ ВШЭ сопоставил первые законодательные инициативы в этой сфере, принятые или обсуждаемые на уровне правительств в Евросоюзе, Китае, России и США.

Весной 2023 г. ученые и представители ИТ-индустрии опубликовали открытое письмо с призывом приостановить хотя бы на полгода обучение систем ИИ мощнее ChatGPT 4, чтобы за это время разработать соответствующие протоколы безопасности. В мае Сэм Альтман, глава компании OpenAI, разработавшей ChatGPT, заявил о необходимости создать специальное агентство для лицензирования систем ИИ, чтобы снизить риски использования технологии. Тогда же стали появляться новости о разработке и принятии законов в области регулирования ИИ различными странами. На данный момент более всего проработаны три законодательных инициативы (табл. 1).

Таблица 1. Сравнение международных инициатив в области регулирования ИИ

Название инициативы	Страна / объединение стран	Дата выхода	Основные положения
Закон об ИИ (AI Act)	Евросоюз	В процессе согласования	• Градация систем ИИ по уровню риска • Запрет вредоносных систем ИИ • Регистрация высокорисковых систем ИИ в европейской базе данных до ввода в эксплуатацию • Создание Управления по ИИ для развития стандартов и тестирования • Установление штрафов за нарушение правил • Создание регуляторной песочницы
Указ о безопасном, надежном и заслуживающем доверия ИИ (Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence)	США	30 октября 2023	• Необходимость сообщать правительству США информацию о результатах тестирования безопасности систем ИИ • Разработка стандартов • Повышение безопасности персональных данных • Запуск Национального ресурса исследований в области ИИ (National AI Research Resource) • Усиление международного сотрудничества
Временные меры по управлению генеративными системами искусственного интеллекта	Китай	10 июля 2023	• Соответствие систем ИИ социалистическим ценностям • Ответственность разработчиков за генерируемый контент • Сотрудничество между различными организациями в сфере ИИ • Укрепление международного сотрудничества • Создание механизма подачи обратной связи

Источник: ИСИЭЗ НИУ ВШЭ по данным Закона ЕС об Искусственном интеллекте (2023), Указа о безопасном, надежном и заслуживающем доверия ИИ (США, 2023), Правил регулирования генеративного ИИ (Китай, 2023).

Европа: градация рисков ИИ-систем

В декабре 2023 г. Европарламент и Евросовет после трехдневного обсуждения согласовали положения Закона об искусственном интеллекте (AI Act). Документ призван защитить от высокорискового искусственного интеллекта (ИИ) гражданские права и демократию, обеспечить верховенство закона и экологическую устойчивость, стимулировать инновации и вывести Европу в лидеры в области ИИ. Еврокомиссар по вопросам внутренней торговли и услуг Тьерри Бретон назвал декабрьское согласование положений «историческим», заявив, что ЕС стал «первым континентом, установившим четкие правила в сфере ИИ».

Подготовка европейского законопроекта началась в 2020 г., его первая версия была опубликована в апреле 2021 г. Законопроекту предстоит еще пройти одобрение государств – членов ЕС, после чего будет опубликована финальная версия документа. Закон вступит в силу не ранее 2025 г.

В основе законопроекта лежит риск-ориентированный подход: документ подразделяет системы ИИ на запрещенные вредоносные ИИ-практики[1]; высокорисковые системы ИИ, на которые как раз и распространяются большинство регуляторных мер[2]; системы ограниченного риска, разработчики которых должны соблюдать требования обеспечения прозрачности процессов; системы с низким или минимальным риском, на которые не налагаются никакие ограничения.

В отношении высокорисковых систем ИИ вводятся следующие правила. Для обеспечения прозрачности компании-разработчики, а также государственные организации, использующие в своей работе решения на базе ИИ, будут обязаны зарегистрироваться в базе данных ЕС по высокорисковым системам ИИ, управляемой Еврокомиссией. Причем разработчики обязаны регистрировать свои продукты в общеевропейской базе данных еще до вывода их на рынок или ввода в эксплуатацию. Высокорисковые системы ИИ должны соответствовать требованиям по риск-менеджменту, тестированию, технической надежности, обучению и управлению данными, прозрачности, кибербезопасности и управляемости людьми. Поставщикам за пределами Евросоюза потребуется уполномоченный представитель в ЕС, который обеспечит оценку соответствия, создаст систему постмаркетингового мониторинга и, при необходимости, предпримет корректирующие действия.

Согласно законопроекту, для систем ИИ, используемых для биометрической идентификации, потребуется оценка со стороны специализированного органа. Особые правила вводятся и для так называемых фундаментальных моделей – систем на основе ИИ, способных компетентно выполнять широкий спектр задач: создание видео, текста, изображений, программного кода. К этой группе относятся и системы генеративного ИИ (одной из них является ChatGPT): их разработчики будут обязаны сообщить, какие объекты авторского права использовались для обучения модели.

Штрафы за нарушения Закона об ИИ устанавливаются на уровне 35 млн евро (3.4 млрд руб.[3]) или 7% годового оборота компании – за нарушение запрета на применение ИИ; 15 млн евро (1.5 млрд руб.) или 3% оборота – за нарушение обязательств по закону об ИИ; 7.5 млн евро (731.6 млн руб.) или 1.5% – за предоставление неверной информации.

В рамках Еврокомиссии будет создано Управление по ИИ (AI Office), которое будет отслеживать передовые модели ИИ, содействовать развитию стандартов и практики тестирования, а также обеспечивать соблюдение правил во всех странах – членах ЕС. Группа независимых экспертов будет консультировать Управление по ИИ по вопросам классификации новых фундаментальных моделей и возможных рисков их применения. Для стимулирования инноваций, создания среды для разработки и тестирования ИИ-решений в реальных условиях организуются регуляторные песочницы.

США: стандарты и безопасность данных

В конце октября 2023 г. президент США Джо Байден подписал Указ о безопасном, надежном и заслуживающем доверия ИИ. Как и европейский законопроект, американский документ требует от создателей систем ИИ прозрачности процессов: они обязаны делиться с американским правительством данными о результатах тестирования безопасности и другой критически важной информацией до того, как разработка выйдет на рынок.

[1] Включают в себя манипулятивные поведенческие техники; системы, эксплуатирующие уязвимые категории населения; системы, используемые для составления социального рейтинга; системы биометрической идентификации в режиме реального времени в общественных местах, используемые в правоохранительных целях, за исключением ограниченного числа случаев (например, для предотвращения теракта или поиска жертв или в рамках расследования тяжких преступлений).

[2] К ним относятся системы, используемые в целях безопасности в продукции, подпадающей под законодательство ЕС о здоровье и безопасности (игрушки, авиация, автомобили, медицинские приборы), а также системы, используемые в восьми категориях: биометрическая идентификация, управление критической инфраструктурой, образование, трудоустройство и управление персоналом, доступ и пользование жизненно важными услугами и благами, правоохранительная деятельность, управление миграцией и пограничным контролем, судебные и демократические процессы.

[3] Суммы в рублях рассчитаны по курсу ЦБ РФ на 11.01.2024, равному 97.5401 за евро.

Для повышения безопасности использования технологий на базе ИИ Национальный институт стандартов и технологий разработает стандарты, которым должны соответствовать эти системы. Министерство торговли создаст руководство по маркировке контента, созданного при помощи ИИ, во избежание распространения такого, который сгенерирован в целях дезинформации.

Особое внимание в американском указе уделяется повышению безопасности персональных данных. Разработками технологий сохранения их конфиденциальности будет заниматься Координационная сеть исследований (объем финансирования в указе не уточняется; проекты, связанные с внедрением передовых технологий сохранения конфиденциальности данных федеральных агентств США, будет поддерживать Национальный научный фонд).

США намерены активизировать исследования в области ИИ посредством Национального ресурса исследований в области ИИ (National AI Research Resource) – пилотного проекта, который предоставит ученым и студентам доступ к данным в сфере ИИ. Кроме того, планируется увеличить число грантов на изучение ИИ в таких областях, как здравоохранение и изменение климата.

США усиливают международное сотрудничество в области ИИ. Так, буквально на следующий день после подписания Джо Байденом рассматриваемого указа вице-президент Камала Харрис приняла участие в первом международном саммите по безопасности ИИ, организованном Великобританией. По его итогам была принята Декларация Блетчли[4], фиксирующая обязательства государств[5] регулировать развитие технологий ИИ.

Китай: особые правила для генеративного ИИ

Принятые в КНР летом 2023 г. Временные меры по управлению генеративными системами искусственного интеллекта являются одними из первых в мире в данной области. Согласно документу, разработчики генеративного ИИ отвечают за весь генерируемый контент. Помимо предписаний улучшать его точность и надежность, в целом повышать прозрачность услуг, разработчики обязаны не создавать контент, подрывающий социалистические ценности или подстрекающий к свержению государственного строя, защищать персональные данные пользователей и соблюдать права на интеллектуальную собственность и частную жизнь. Правилами устанавливается необходимость предотвращать дискриминацию по какому-либо признаку и чрезмерную зависимость несовершеннолетних от использования генеративного ИИ. Разработчики обязаны разъяснять, как именно применяется контент, предлагать понятный механизм подачи и обработки жалоб и обратной связи. Правила поощряют сотрудничество в сфере ИИ между предприятиями, университетами, научно-исследовательскими институтами и государственными учреждениями, а также участие представителей КНР в разработке международных правил, связанных с генеративным ИИ. Сотрудничество с другими странами должно строиться на равноправной и взаимовыгодной основе.

Опыт России

В России также готовится законопроект о регулировании в сфере ИИ. Как сообщил в апреле 2023 г. зампред комитета Госдумы по информационной политике Антон Горелкин, особое внимание в документе будет уделяться защите рынка труда от негативного влияния ИИ. Законопроект должен определить ответственность за разработку систем ИИ и создаваемый с их помощью контент и исключить использование технологии мошенниками.

На ускорение темпов развития ИИ нацелена и подготовка стандартов в этой сфере. В июне 2023 г. стало известно, что Минцифры России создает стандарт персональных данных для систем и сервисов, использующих средства ИИ, а Росстандарт открыл доступ к более чем 60 стандартам в сфере ИИ, подготовленным Техническим комитетом по стандартизации № 164 «Искусственный интеллект».

Свои рекомендации относительно применения ИИ на финансовом рынке в ноябре выпустил ЦБ РФ. Придерживаясь в этом вопросе риск-ориентированного подхода, регулятор планирует, например, участвовать в отслеживании оборота обезличенных персональных данных, выработке подходов к распределению ответственности за генерируемый контент и использованию разработчиками сторонней инфраструктуры обработки данных.

В середине декабря правительство внесло в Госдуму законопроект об изменениях в Федеральном законе «Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации». Инициатором нововведений выступило Минэкономразвития России, которое еще

[4] По названию усадьбы Блетчли-парк, где проходил саммит, а в годы Второй мировой войны находилась правительственная школа кодов и шифров и работала команда математика Алана Тьюринга.

[5] Декларацию подписали Австралия, Бразилия, Великобритания, Германия, Евросоюз, Израиль, Индия, Индонезия, Ирландия, Испания, Италия, Канада, Кения, Китай, Республика Корея, Нигерия, Нидерланды, ОАЭ, Руанда, Саудовская Аравия, Сингапур, США, Турция, Украина, Филиппины, Франция, Чили, Швейцария, Япония.

весной 2023 г. опубликовало поправки к закону об антикоррупционной экспертизе. Согласно предложенным изменениям, компании – участники экспериментальных правовых режимов (ЭПР) должны будут вести, во-первых, учет технологий, созданных с применением ИИ; во-вторых, реестр лиц, с которыми возникли правоотношения (в частности, в нем будут фиксироваться сведения о лицах, ответственных за причинение вреда в результате использования технологий, созданных с применением ИИ). Участники ЭПР будут обязаны страховать гражданскую ответственность за вред, который может возникнуть в результате использования ИИ. Выявление причин и обстоятельств, вызвавших такие последствия, возлагается на специально созданную комиссию, в составе которой будут представители уполномоченного и регулирующего органов, предпринимательского сообщества и иные лица. Отчеты комиссии о причинах и обстоятельствах произошедшего, а также о мерах по возмещению ущерба будут публиковаться в интернете.

Резюме

Развитие ИИ открывает огромные возможности в самых разных областях жизнедеятельности человека и подводит к необходимости решить нетривиальную дилемму. С одной стороны, стремительно усиливается конкуренция за лидерство в разработке ИИ; с другой – свободное развитие технологии несет в себе серьезные социальные и экономические риски, предотвращение которых требует четкого регулирования. Достигнуть лидерства удастся тем странам, которые смогут в том числе на законодательном уровне найти оптимальный баланс между поддержкой ИИ и его ограничением.

Источники: официальные документы Великобритании, ЕС, Китая, РФ, США, 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 подготовил: **З.А. Мамедьяров**