

Аналитический обзор



**ЗАРУБЕЖНЫЕ ПОДХОДЫ К РЕГУЛИРОВАНИЮ
БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ (LLM)**



**CENTER
FOR GLOBAL
IT-COOPERATION**

Центр глобальной ИТ-кооперации

Москва 2026

Оглавление

ВВЕДЕНИЕ	6
1. ЦИФРОВОЙ СУВЕРЕНИТЕТ: ПОДХОДЫ К ОПРЕДЕЛЕНИЮ	7
1.1 Европейский подход (стратегическая автономия и защита прав)	8
1.2 Подход, ориентированный на государственный контроль и безопасность	8
1.3 Аналитические и нормативные документы, стандартизация.....	9
2. РЕГУЛИРОВАНИЕ LLM: СЛОЖНОСТИ ГАРМОНИЗАЦИИ	10
2.1 Обсуждаемые и тестируемые магистральные пути решения	12
2.2 Примеры регулирования вредоносного контента.....	12
2.3 Ключевые законодательные подходы	13
3. РЕГУЛИРОВАНИЕ LLM В НЕКОТОРЫХ РЕГИОНАЛЬНЫХ И НАЦИОНАЛЬНЫХ ЗАКОНАХ	15
3.1 Азиатско-Тихоокеанский регион: диверсификация подходов	15
3.2 Ближний Восток: Регулирование как часть экономической диверсификации	16
3.3 Великобритания, Франция	16
3.4 Американский континент	17
4. СТАНДАРТИЗАЦИЯ В СФЕРЕ ПРИМЕНЕНИЯ LLM	21
5. НАУЧНЫЕ РАБОТЫ ПО ПРОБЛЕМЕ РЕГУЛИРОВАНИЯ LLM	23
6. НЕКОТОРЫЕ ПРАКТИКИ И РЕЖИМЫ ДОСТУПА К LLM	24
7. ОБЩИЕ ВЫВОДЫ	25

Зарубежные подходы к регулированию больших языковых моделей (LLM), Аналитический обзор, Игнатъев А.Г., Центр глобальной ИТ-кооперации, 26 С, 2026

Краткая аннотация: В обзоре представлены общие подходы стран к созданию правовой базы в области регулирования искусственного интеллекта с акцентом на законодательные рамки для нейросетевых языковых моделей (LLM). Обзор не претендует на полное и детальное отражение международных практик в данной области, а отражает лишь наиболее значимые тенденции и инициативы стран, которые уже приступили к выработке правовых норм для LLM применительно к различным этапам жизненного цикла таких систем в различных средах применения.

С учетом влияния LLM на весь комплекс вопросов, связанных с информационной безопасностью, в обзоре делается попытка отразить смысл и наполнение понятия «цифровой суверенитет», обозначить связь и взаимопроникновение законодательных норм в сфере LLM с другими областями регулирования цифровой среды. Обозначены ключевые факторы, влияющие на дальнейшее развитие процесса разработки соответствующих инструментов и стандартов.

Обзор не включает в себя анализ российских национальных законов, стратегий, программ и этических кодексов в сфере ИИ.

Материал может быть полезен в качестве справочного материала профильным организациям, научно-исследовательским центрам, учебно-образовательным заведениям и различным площадкам развития в области цифровых технологий.

Ключевые слова: искусственный интеллект (ИИ), Большие языковые модели (БЯМ), регулирование в области ИИ, законы, инструменты, стандарты, декларации, соглашения, мягкое право в области ИИ, технические стандарты и стандартизация в области ИИ.



АНО «Центр компетенций по глобальной ИТ-кооперации» создан в 2020 году для экспертного изучения вопросов международного сотрудничества в сфере информационных технологий (ИТ), укрепления позиций России в глобальной ИТ-кооперации, в частности, продвижения новых подходов к многостороннему и равноправному управлению Интернетом на основе обеспечения безопасности и уважения национального суверенитета.

CGITC является членом Сектора развития электросвязи (ITU-D) Международного союза электросвязи, с 2021 года Центр представляет Россию на ежегодном международном Форуме ООН по управлению Интернетом (IGF).

Центр проводит исследования и реализует проекты в области международного сотрудничества в сфере цифровой экономики (развитие всего стека цифровых технологий, политика регулирования, управление Интернетом, научно-техническая кооперация, цифровая грамотность и др.), оказывает содействие продвижению экспорта продуктов и услуг в сфере ИКТ.

Во взаимодействии с международным сообществом и при поддержке заинтересованных специалистов в России CGITC на регулярной основе проводит ряд научных и экспертных круглых столов, конференций и вебинаров.

CGITC является соорганизатором ежегодного Российского форума по управлению Интернетом, ключевым организатором Молодежного российского форума по управлению Интернетом, в 2022 и 2023 годах - участник проекта Think20 исследовательской сети G20.

Правила использования Обзора

Настоящий аналитический обзор «Зарубежные подходы к регулированию больших языковых моделей (LLM)» (далее – «Обзор») подготовлен АНО «Центр глобальной ИТ-кооперации».

Информация, приведенная в Обзоре, подпадает под действие законодательства об авторских правах Российской Федерации. Исключительные права на Обзор принадлежат АНО «Центр глобальной ИТ-кооперации» (далее – «Правообладатель»). Права на обзор могут быть переданы другой организации по отдельному соглашению.

Обзор может использоваться в целях ознакомления. Допускается размещение активных ссылок на Обзор в других информационных источниках без непосредственного копирования его содержания. При любом использовании Обзора активная ссылка на источник обязательна.

Частичное или полное воспроизведение и распространение, а также любое коммерческое использование Обзора запрещено без письменного разрешения правообладателя, а также без ссылки на авторов Обзора.

Приступая к ознакомлению с Обзором, вы подтверждаете свое согласие с изложенными ниже условиями:

- Правообладатель не принимает на себя обязательств или ответственности за использование информации, содержащейся в Обзоре.
- Обзор носит исключительно информационный характер и составлен на основе публичных (открытых) источников, признанных надежными, однако правообладатель не несет ответственности за точность приведенных данных.
- Выводы, представленные в Обзоре, носят исключительно информационный характер и основаны на информации, полученной из открытых источников, указанных в соответствующих ссылках.
- Обзор не является юридическим заключением по вопросам, рассмотренным в нем. Правообладатель не несет ответственности за решения, принятые на основании представленных в Обзоре данных.
- Обзор включает в себя ссылки на сторонние веб-сайты, находящиеся вне контроля правообладателя. Правообладатель не несет ответственности за содержание этих ссылок; такая ответственность во всех случаях возлагается на соответствующего провайдера, либо оператора этих сторонних веб-сайтов. Правообладатель не несет ответственности за доступ к этим веб-сайтам и их содержанию.

ВВЕДЕНИЕ

Большие языковые модели (large language model, LLM) на основе технологии искусственного интеллекта (ИИ) являются на сегодня наиболее востребованными во всех областях хозяйственной и повседневной жизни, активно развиваются и модернизируются.

Общепринятого универсального определения LLM не выработано. Основные существующие формулировки определяют Большую языковую модель (рус. аббревиатура - БЯМ) как одну из систем на основе искусственного интеллекта (относят к т. н. генеративному ИИ), созданную с помощью алгоритмов глубокого обучения, для распознавания, генерации, перевода и/или обобщения огромных объемов письменного человеческого языка и текстовых данных. Такие модели используют наиболее передовые решения в области обработки естественного языка (NLP).

В ядро глобальной экосистемы LLM входят следующие наиболее распространенные зарубежные модели (список не является исчерпывающим):

- GPT (Generative Pre-trained Transformer) от OpenAI – флагманская серия моделей, включающая ChatGPT различных модификаций;
- Gemini от Google – мультимодальное семейство моделей, ориентированное на решение сложных задач. Gemini 2.5 Pro считается одной из лидирующих моделей в рейтингах 2025 года;
- Claude от Anthropic – позиционируется как одна из лучших для написания кода и длительных задач рефакторинга. Claude Opus 4 и Claude Sonnet 4 активно используются разработчиками;
- Llama (Large Language Model Meta AI) от Meta¹ – семейство открытых моделей, которые могут быть дообучены и внедрены под конкретную задачу. Llama 3 и более поздние версии популярны в исследовательской и коммерческой среде по всему миру;
- Grok 4.1 - обновленная версия ИИ-модели от xAI Илона Маска (ноябрь 2025), предполагает по своему замыслу высокий эмоциональный интеллект, улучшенное понимание намерений и снижение галлюцинаций; существует две версии: быстрая (Fast) для диалогов и рассуждающая (Thinking) для сложных задач, сохраняя доступ через платформу X²;
- BLOOM (BigScience Large Open-science Open-access Multilingual Language Model) – открытая многоязычная модель, созданная международным коллективом исследователей. Она доступна через Hugging Face и предназначена для поддержки множества языков;
- ERNIE (Enhanced Representation through kNowledge IntEgration) от китайского Baidu – серия моделей, усиленных знаниями, является основным продуктом Baidu в области генеративного ИИ;

¹ Meta Platforms Inc. (владелец Facebook и Instagram) — организация признана экстремистской, её деятельность запрещена на территории России.

² Платформа X (ранее Twitter) заблокирована на территории России, доступ к ней ограничен Роскомнадзором с начала марта 2022 г.

- Tongyi Qianwen (Qwen) от Alibaba Cloud – семейство больших языковых и мультимодальных моделей, которые Alibaba предоставляет open-source сообществу;

- DeepSeek – китайская модель, которая в 2025 году отмечалась за глубину рассуждений и высокие результаты в бенчмарках.

- Perplexity - мультимодельная система с собственной моделью Sonar (лучший по точности фактов), построенной на базе Llama 3.1 70B, которая также использует GPT 5.2 (для сложных рассуждений), Claude Sonnet 4.5 (для логической ясности), Gemini 3 Pro (для мультимодальных задач), Grok 4.1 (для режима в реальном времени) и Kimi K2 (для конфиденциальных запросов).

Транснациональный характер перечисленных моделей проявляется не только в географической доступности, но и в многоязычной поддержке, открытых лицензиях для многих проектов и интеграции в международные облачные платформы. Несмотря на возможные ограничения доступа к некоторым проприетарным моделям в отдельных странах, эти LLM остаются ключевыми игроками на мировом рынке ИИ, включая Россию. Широко используются во всём мире через API и облачные сервисы, а также через локальные внедрения.

Сегодня страны активно развивают законодательное регулирование ИИ, в т. ч. пытаются распространять его на большие языковые модели (LLM).

Основные цели такого регулирования включают соблюдение национальных законов о данных, защиту традиционных ценностей и предотвращение потенциального вреда.

Все более отчетливо прослеживается, что использование LLM в целом ряде случаев могут вызывать сложности в соблюдении норм национального законодательства. Кроме того, режимы использования моделей и сами их решения могут не соответствовать традиционным ценностям и общепринятым этическим нормам, входить в противоречие с положениями по защите интеллектуальной собственности.

Вопросы, связанные с тем, как устранить подобного рода конфликты в широкомасштабной практике использования моделей LLM приобретают все большее значение в законодательной деятельности по регулированию цифровых технологий.

Комплекс вопросов по регулированию LLM тесно связан с защитой национального **цифрового суверенитета**.

1. ЦИФРОВОЙ СУВЕРЕНИТЕТ: ПОДХОДЫ К ОПРЕДЕЛЕНИЮ

В контексте темы регулирования LLM уместно представить информацию о подходах к данному термину - «цифровой суверенитет». Этот вопрос является довольно сложным и затрагивает как политико-правовую сферу, так и техническую стандартизацию.

Единого юридически закреплённого определения «цифровой суверенитет» не существует.

Термин «цифровой суверенитет» (Digital Sovereignty) на сегодняшний день не имеет универсального, общепризнанного определения в международном праве или в ключевых межгосударственных документах. Это стратегическая и политическая концепция, наполняемая разным содержанием в зависимости от субъекта (государство, блок государств, корпорация). Однако существуют документы и аналитические работы, которые формулируют его суть.

Ниже приведены ссылки на наиболее авторитетные источники, отражающие разные подходы.

1.1 Европейский подход (стратегическая автономия и защита прав)

Коммюнике Европейской Комиссии «Формирование цифрового будущего Европы» (Shaping Europe's Digital Future, 2020). В этом документе впервые на уровне институтов ЕС четко провозглашается цель достижения «цифрового суверенитета» как способности Европы действовать независимо в цифровом мире³.

Декларация о цифровом суверенитете ЕС (Declaration on Digital Sovereignty, 2020)⁴. Подписанная лидерами ЕС, эта декларация закрепляет политическую приверженность концепции, связывая ее с защитой данных, инфраструктурой и технологическим лидерством

Стратегия ЕС в области данных и коммюнике Комиссии «Формирование цифрового будущего Европы»⁵.

Определение Европейского Союза можно считать своеобразным эталоном для демократических стран: **цифровой суверенитет** — это «способность Европы действовать самостоятельно в цифровом мире».

Это, в частности, подразумевает:

- технологическую независимость: контроль над критически важными цифровыми инфраструктурами, полупроводниками, платформами;
- нормативную самостоятельность: возможность устанавливать свои правила (как GDPR, AI Act, DSA), которые становятся глобальными стандартами;
- защиту данных и прав граждан: обеспечение того, чтобы данные европейцев обрабатывались в соответствии с европейскими ценностями и законами.

Документ Gaia-X "Technical Architecture – The Concept of Digital Sovereignty Explained" (2021)⁶ (европейская инициатива по созданию облачной инфраструктуры) дает техническое определение цифрового суверенитета с точки зрения контроля пользователя над данными и процессами. Так, в частности, в документе указано, что суверенность обмена данными обеспечивается механизмами контроля за использованием и всеобъемлющей концепцией безопасности. Кроме того, разрабатываются соответствующие стандарты для обеспечения функциональной совместимости обмена данными.

1.2 Подход, ориентированный на государственный контроль и безопасность

Китай: «Закон КНР о кибербезопасности» (2017). Хотя термин «суверенитет» прямо используется в преамбуле в контексте «киберпространственного суверенитета», весь закон является правовым отражением этой концепции через контроль над данными, инфраструктурой

³ https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/shaping-europe-digital-future_en

⁴ <https://www.consilium.europa.eu/media/48014/declaration-on-digital-sovereignty-en.pdf>

⁵ <https://eufordigital.eu/ru/library/shaping-europes-digital-future/>

⁶ https://www.bundeswirtschaftsministerium.de/Redaktion/EN/Publikationen/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=1

и контентом (см. ссылку на официальный текст на китайском языке⁷, а также перевод на английский⁸).

1.3 Аналитические и нормативные документы, стандартизация

OECD Digital Economy Paper «Data Sovereignty: International and Regulatory Perspectives» (2023)⁹. Хотя основной фокус делается на данных, документ дает определение суверенитета в цифровой среде и анализирует его проявления.

Прямого термина «Digital Sovereignty» в ISO/IEC на данный момент нет. Однако существует фундаментальная концепция, которая является его технико-правовой основой — это «Sovereignty over data» или, более формально, принципы управления данными, обеспечивающие контроль.

Подкомитет ISO/IEC JTC 1/SC 38 (Cloud Computing and Distributed Platforms)¹⁰ разрабатывает стандарты для различных аспектов облачных вычислений, которые критичны для суверенитета: локализация данных, прозрачность цепочек обработки и т.п.

ISO/IEC 19944-1:2020 (Cloud computing and distributed platforms — Data flow, data categories and data use)¹¹ определяет где и как обрабатываются данные, что напрямую связано с возможностью осуществления их контроля - одна из задач для обеспечения цифрового суверенитета в области данных.

Стандарты ISO/IEC раскрывают концепцию цифрового суверенитета через конкретные технические и организационные меры: криптография, управление доступом, политики хранения, требования к контрактам с облачными провайдерами. Они отвечают на вопрос «как» обеспечить соответствующий контроль при реализации операций в цифровой среде.

Существующие ISO/IEC стандарты по кибербезопасности (ISO/IEC 27000 серия), облачным вычислениям (проекты Подкомитета SC 38) и потокам данных (ISO/IEC 19944) создают технический фундамент для реализации требований суверенитета.

В России Федеральный закон № 187-ФЗ «О безопасности критической информационной инфраструктуры РФ»¹² и закон № 90-ФЗ "О внесении изменений в Федеральный закон "О связи" рассматривают цифровой суверенитет как возможность обезопасить национальный сегмент Интернета и обеспечить контроль над критической инфраструктурой. Россия фокусируется на концепции технологического суверенитета.

⁷ http://www.cac.gov.cn/2016-11/07/c_1119867116.htm

⁸ <https://digichina.stanford.edu/work/translation-cybersecurity-law-of-the-peoples-republic-of-china/>

⁹ https://www.oecd-ilibrary.org/science-and-technology/data-sovereignty-international-and-regulatory-perspectives_b7f0f2c1-en

¹⁰ <https://www.iso.org/committee/601355.html>

¹¹ <https://www.iso.org/standard/79573.html>

¹² https://www.consultant.ru/document/cons_doc_LAW_220885/

Отдельный закон об ИИ пока не принят, но действует Концепция развития регулирования ИИ до 2024 года. Определенные ограничения на LLM вытекают из положений о локализации данных (152-ФЗ) и 187-ФЗ.

Анализ вышеперечисленных подходов показывает, что для системного понимания термина «цифровой суверенитет» необходимо изучать его, как минимум, в трех контекстах:

- правовые подходы и документы конкретного государства или региона;
- развивающиеся технические стандарты в области кибербезопасности, данных и вычислений, в т.ч. стандарты в ISO/IEC;
- отраслевые практики корпоративного суверенитета, в т. ч. обеспечение суверенитета над данными.

2. РЕГУЛИРОВАНИЕ LLM: СЛОЖНОСТИ ГАРМОНИЗАЦИИ

Специалисты отмечают, что важной проблемой является фундаментальное противоречие между транснациональной, децентрализованной природой больших языковых моделей и территориально ограниченной, фрагментированной (сложно влиять на все аспекты) системой национального регулирования. Это в свою очередь порождает проблему, которую можно условно обозначить как «проблема трех несоответствий».

а) Проблема несоответствия подходов в различных юрисдикциях (jurisdictional mismatch).

Речь идет о том, что физическая инфраструктура, разработчики, данные для обучения и конечные пользователи LLM находятся в разных странах, в разных юрисдикциях, что делает невозможным однозначное определение и имплементацию правовых норм.

Так, например модель LLM может быть разработана и обучена в США на данных, собранных по всему миру, размещена на серверах в Ирландии, напрямую или через специальные сервисы доступна в России.

Ввиду этого возникает конфликт подходов, т. е. неопределенность в применении принципов:

- принцип территориальности: государство регулирует деятельность на своей территории (например, китайские законы об ИИ, российский закон о "суверенном интернете" и т.п.). Однако данный принцип сложно применять, когда речь идет о доступе к платформам через VPN или API.
- принцип экстерриториальности: государство пытается регулировать деятельность зарубежных платформ и продуктов, если они затрагивают интересы страны или ее граждан (как GDPR через механизм "переноса данных" или американские санкции в сфере ИИ). Это приводит к конфликту норм в различных юрисдикциях и возникновением соответствующих противоречий между заинтересованными участниками.

Доклад UNIDIR (Институт ООН по исследованию проблем разоружения) "Аспекты управления ИИ, связанные с безопасностью" (2023)¹³ указывает, что трансграничный характер систем ИИ

¹³ <https://unidir.org/publication/artificial-intelligence-beyond-weapons-application-and-impact-of-ai-in-the-military-domain/>

создает проблемы для правоприменения, основанного на нормах определенной юрисдикции. Анализ Брукингского института "Правоприменение GDPR за пределами ЕС" (G. Bradford, 2020) описывает сложности экстерриториального применения правил.

В статье *The Wild West: A Mosaic of International Approaches to Global AI Regulation* отмечается, что «...различия в управлении ИИ являются не просто отражением национальных приоритетов, но и результатом различных политических, экономических и социальных контекстов, в которых эти технологии разрабатываются и внедряются»¹⁴.

б) Проблема архитектурного несоответствия (architectural mismatch)

Техническая архитектура LLM (непрозрачность, стохастичность, масштаб, охват и др.) не всегда вписывается, а зачастую входит в противоречие с базовыми требованиями современных режимов регулирования.

Это обусловлено невозможностью выполнения некоторых ключевых правовых требований, таких как:

- право на объяснение (GDPR, ст. 22): как объяснить конкретное решение "черного ящика" с миллиардами параметров? Это подрывает принцип справедливой обработки;
- право на удаление данных ("право на забвение", GDPR, ст. 17): машинное "забывание" (machine unlearning) для LLM является технологически нерешенной задачей; удалить данные из обученной модели, не нарушив/изменив ее алгоритм, практически невозможно;
- право на исправление неточных данных (CPRA)¹⁵: если модель сгенерировала о человеке ложный факт и запомнила его - исправить такую запись достаточно сложно, это нетривиальная инженерная задача.

Таким образом возникают проблемы оценки/аудита непрозрачных моделей, в т. ч. оценки справедливости решений моделей¹⁶. При этом ведущие разработчики LLM не раскрывают критически важную (в т. ч. и для регуляторов) информацию.

в) Проблема нормативного несоответствия (Regulatory Mismatch)

Скорость технологического развития LLM на порядок превышает скорость законотворческого процесса (разработка столь сложных законов занимает, как правило, годы). Передовые модели с совершенно новыми возможностями (мульти-modalность, мультиагентность и др.) появляются быстрее, чем регуляторы успевают адаптировать или принять подзаконные акты к уже существующим технологиям.

Фрагментацию норм и различные акценты в законодательной политике можно проследить на следующих примерах:

- ЕС (AI Act): акцент на оценке риска, фундаментальных правах, запрете отдельных приложений;

¹⁴ <https://www.fordhamilj.org/iljonline/k73pddfmra65ecd-fxc5k-bdazp-npp5j-l8be9-kbndm-tlwl-f-2pnhz-hpjc3-2bxkm-9egc4-4acpz-ez2w6>

¹⁵ CPRA (Закон Калифорнии о правах на конфиденциальность) дает потребителям право требовать от компаний исправления неточных персональных данных. Если данные неточны, компания должна обновить их во всех своих системах, включая данные, переданные третьим лицам

¹⁶ <https://www.nature.com/articles/s41598-024-68651-w>

- США (исполнительные указы, законы штатов): акцент на инновациях, отраслевом регулировании и национальной безопасности;
- Китай (требования для генеративного ИИ): акцент на безопасности контента, идеологическом соответствии и контроле данных.

При этом попытка соблюсти все регуляторные режимы одновременно ведет либо к блокированию части информационной среды (Интернета и платформ), либо к перемещению разработок в юрисдикции с самым мягким регулированием.

На основе изложенного можно констатировать, что ключевая проблема регулирования транснациональных LLM — это системный кризис управления, обусловленный сложностью, при котором устаревшие, территориально привязанные правовые инструменты пытаются контролировать принципиально новую, трансграничную, самообучающуюся и непрозрачную технологическую силу.

2.1 Обсуждаемые и тестируемые магистральные пути решения

2.1.1 Гармонизация правил на уровне международных организаций (ОЭСР, Совет Европы, G20).

2.1.2 Регулирование через "шлюзы" (gatekeepers): контроль не за самой технологией, а за точками её входа в пользовательскую среду (магазины приложений, облачные платформы, платежные системы), как в DSA ЕС.

2.1.3 Сдвиг к отраслевому регулированию по результатам оценки рисков (risk-based): фокусировка не на архитектуре LLM, а на контексте её применения (медицина, финансы, правосудие и др.) с соответствующим уровнем надзора.

2.1.4 Развитие технических стандартов (ISO/IEC, IEEE): создание единых технических протоколов для аудита, тестирования и сертификации, которые могут быть инкорпорированы в национальные законы.

2.2 Примеры регулирования вредоносного контента

- США (Section 230 CDA): платформа (провайдер LLM) в целом не несет ответственности за контент, сгенерированный пользователем/моделью;
- ЕС (DSA): крупные платформы обязаны внедрять меры по снижению рисков и удалять незаконный контент. AI Act добавляет требования к прозрачности для генеративного ИИ;
- Китай: разработчик LLM несет полную ответственность за весь сгенерированный контент и обязан внедрить системы/механизмы для его предварительной цензуры.

С учетом современных инструментов, которые ориентированы на самую широкую аудиторию и географию, компания-разработчик, развертывающая глобальную LLM, должна технически создать три разных версии системы модерации, соответствующих разным юридическим стандартам, и точно идентифицировать юрисдикцию каждого пользователя. Это экономически и технически обременительно.

Для углубленного изучения проблемы, помимо законодательных инструментов ЕС, США и Китая (OECD AI Principles, 2019 и последующие отчеты о их внедрении, AI Act, DSA, стратегия США по ИИ, 2023 и соответствующие исполнительные указы, национальная стратегия Китая по развитию ИИ и "Правила управления генеративным ИИ"), также могут быть изучены аналитические отчеты ведущих think tanks: Brookings Institution, Carnegie Endowment for International Peace, Center for Security and Emerging Technology (CSET) и др.

Для поиска решений по регулированию LLM могут быть также изучены инструменты международных организаций в области управления ИИ¹⁷

2.3 Ключевые законодательные подходы

2.3.1 Регулирование, основанное на рисках (подход ЕС, AI Act)

Это первый в мире комплексный закон об ИИ. Документ классифицирует системы ИИ (включая LLM) по уровням риска: неприемлемый, высокий, ограниченный и минимальный.

LLM общего назначения (как GPT-4, Gemini) подпадают под особые требования к прозрачности (раскрытие, что контент создан ИИ, обязательное информирование пользователей о возможных ошибках);

Системы, признанные высокорисковыми (например, используемые в правосудии, образовании, критической инфраструктуре), подвергаются строгому контролю: оценка соответствия, качество данных, регистрация в базе данных, человеческий надзор;

Запрещаются («неприемлемый риск») системы, которые используют социальный скоринг (как в Китае), манипулируют поведением или эксплуатируют уязвимости определенных групп — что напрямую связано с защитой ценностей и прав граждан.

2.3.2 Защита персональных данных

GDPR в ЕС, применительно к данным, ограничивает работу LLM за счет следующих требований:

- необходимость правового основания для обработки данных (информированное согласие и т.д.);
- возможность удаления персональных данных («право на забвение»), что технически сложно для моделей, уже обучившихся на этих данных;
- ограничения на профилирование, т. е. на автоматизированную обработку данных для анализа личных аспектов, связанных с конкретным лицом;
- прозрачность в использовании данных.

Аналогичные законы есть во многих странах (например, LGPD в Бразилии, CCPA/CPRA в Калифорнии, США). Они создают серьезные барьеры для сбора данных для обучения моделей и их развертывания.

¹⁷ Управление ИИ: Актуальные международные инструменты, Аналитический обзор, Игнатъев А.Г., Курбатова Т.А., Шамраев Р.А., Центр глобальной ИТ-кооперации, 33 С, 2025 <https://cgitc.ru/research/upravlenie-ii-aktualnye-mezhdunarodnye-instrumenty/>

2.3.3 Законы о цифровых услугах и контенте

Digital Services Act (DSA) в ЕС обязывает крупные онлайн-платформы и поисковые системы (через которые часто предоставляется доступ к LLM) бороться с незаконным контентом, защищать пользователей, обеспечивать прозрачность рекомендательных алгоритмов. Это может затрагивать и решения LLM.

Закон о безопасности интернет-информации в Китае требует, чтобы контент в сети, включая генерируемый ИИ, соответствовал социалистическим ценностям, не угрожал национальной безопасности и общественному порядку. Это прямое регулирование в целях защиты государственных ценностей.

2.3.4 Стратегии, ориентированные на национальную безопасность и ценности

Китай ввел одни из самых жестких в мире правила управления генеративным ИИ и глубоким синтезом. Они требуют, чтобы контент ИИ соответствовал социалистическим ценностям, не распространял «подрывную» информацию, проходил обязательную проверку безопасности перед выпуском. Также действуют строгие законы о кибербезопасности и данных, требующие хранения данных на территории Китая.

В России действует Закон о «суверенном интернете» (ФЗ № 90-ФЗ «О внесении изменений в Федеральный закон „О связи“ и Федеральный закон „Об информации, информационных технологиях и о защите информации“»), 2019, а также Стратегия развития ИИ. Акцент делается на технологический суверенитет. Положения о персональных данных (требование локализации) и о «фейковых новостях» могут использоваться для регулирования контента, генерируемого LLM, если он признан вредным или не соответствующим государственной политике.

В США подход более фрагментирован, но регулирование происходит через исполнительные указы (существует требование к разработчикам ИИ о необходимости делиться данными в сфере безопасности с правительством), отраслевые законы (например, HIPAA в здравоохранении) и законы штатов. Акцент делается на конкуренцию, инновации и национальную безопасность, но при этом уделяется внимание и защите гражданских прав.

2.3.5 Подход, основанный на принципах (Великобритания, Япония, Канада)

Эти страны пока избегают жесткого сквозного регулирования, предпочитая гибкие рамки, основанные на принципах (безопасность, прозрачность, подотчетность). Регулирование часто привязано к существующим законам о данных, о конкуренции, о защите прав потребителей. Однако и в них обсуждается необходимость применения более строгих мер.

Следует отметить, что отдельные страны пытаются адаптировать подход ЕС, США и Китая, создавая т. н. гибридные модели, тем не менее такие законы так или иначе имеют один – два направленных акцента из перечисленных выше (2.3.1 – 2.3.5).

3. РЕГУЛИРОВАНИЕ LLM В НЕКОТОРЫХ РЕГИОНАЛЬНЫХ И НАЦИОНАЛЬНЫХ ЗАКОНАХ

Для комплексного изучения проблемы регулирования больших языковых моделей (LLM) в условиях, когда в юрисдикциях формируются свои национальные подходы, возможно будет полезно выйти за рамки наиболее обсуждаемых актов, принятых или готовящихся в ЕС, США и Китае. Глобальная регуляторная картина говорит об отсутствии гармонизации и легких путей нахождения правовых компромиссов. Государства и регионы предлагают принципиально отличные подходы, создавая сложную среду для разработчиков транснациональных систем ИИ.

Ниже представлен обзор некоторых законодательных инициатив, заслуживающих внимания применительно к вопросам внедрения и использования больших языковых моделей на основе ИИ.

3.1 Азиатско-Тихоокеанский регион: диверсификация подходов

В АТР, помимо жёсткой китайской модели, сосуществуют несколько иных регуляторных подходов.

Япония делает ставку на «мягкое» регулирование, основанное на принципах, в рамках своей стратегии Society 5.0¹⁸. Акцент на инновациях и человекоцентричности предполагает создание руководств, а не жестких запретов, что создаёт предсказуемую, но не обременительную среду для разработки LLM.

Сингапур развивает уникальную модель добровольного, но структурированного аудита через свою Модельную рамку управления ИИ и инструмент AI Verify¹⁹. Такой подход, основанный на рекомендациях, позволяет компаниям демонстрировать соответствие стандартам прозрачности и безопасности, что может стать моделью для отраслевой сертификации LLM.

Южная Корея приняла собственный Акт о развитии и распространении ИИ²⁰, который представляет собой гибридную модель. Закон сочетает меры по стимулированию инноваций с оценкой рисков для высокорисковых систем, одновременно находясь в синергии со строгим законом о защите персональных данных (PIPA).

Индия после принятия Закона о цифровых персональных данных (DPDPA, 2023)²¹ заложила основу для развития своего цифрового регулирования. Хотя единого закона об ИИ пока нет, DPDPA с его акцентами на суверенитете данных и локализации формирует критически важные ограничения для развёртывания и использования LLM на индийском рынке.

¹⁸ https://www8.cao.go.jp/cstp/english/society5_0/index.html

¹⁹ <https://aiverifyfoundation.sg/>

²⁰ <https://www.law.go.kr/>

²¹ <https://www.meity.gov.in/data-protection-framework>

3.2 Ближний Восток: Регулирование как часть экономической диверсификации

Страны Персидского залива активно используют регулирование ИИ как инструмент для привлечения инвестиций и построения пост-нефтяной экономики.

ОАЭ стали пионерами в регионе, утвердив рамочный пакет документов, регулирующих ИИ²², в том числе включая The UAE Charter for the Development and Use of Artificial Intelligence²³. Кроме того, в стране создан профильный регулятор ADIAIC²⁴ (Международный арбитражный центр Абу-Даби, ArbitrateAD выступает в качестве нейтрального и беспристрастного форума по разрешению споров для коммерческих и государственных организаций). В рамках законодательных мер предусмотрен механизм лицензирования для высокорисковых систем ИИ, прямо затрагивая профессиональные и корпоративные применения LLM.

Саудовская Аравия в рамках Vision 2030 фокусируется на масштабных инвестициях и развитии инфраструктуры, что отражено в её Национальной стратегии в области данных и ИИ²⁵. Регуляторная среда здесь нацелена на облегчение внедрения и тестирования технологий, включая LLM, что создаёт зону с относительно низкими первоначальными барьерами, но с потенциальным усилением контроля по мере развития экосистемы.

3.3 Великобритания, Франция

Несмотря на общеевропейский AI Act, национальные стратегии в европейских странах представляют важность и заслуживают внимания.

Великобритания после выхода из ЕС избрала направление развития ИИ, изложенное в Белой книге по регулированию ИИ²⁶ (в т. ч. включает регулирование в сфере оценки воздействия - UK Artificial Intelligence Regulation Impact Assessment²⁷). Вместо создания нового регулятора, UK усиливает полномочия существующих органов (ICO, CMA) на основе пяти адаптируемых принципов. Это создаёт менее жесткую, но потенциально более гибкую среду для LLM в разных секторах экономики.

Франция, являясь активным участником европейского регулирования, параллельно продвигает повестку цифрового суверенитета. Это выражается в прямой государственной поддержке

²² Key Pillars of the UAE AI Regulatory Framework:

UAE Charter for the Development and Use of AI (2024): A 12-principle framework covering human oversight, safety, transparency, accountability, and fairness.

AI Principles and Ethics: An 8-point, "living document" ensuring AI operates without bias, respects privacy, and ensures sustainability.

Data Protection & Privacy: Federal Decree-Law No. 45 of 2021 mandates lawful processing of personal data, which is essential for AI systems.

Sectoral Regulations: The Dubai International Financial Centre (DIFC) has specific amendments addressing AI accountability and governance.

Institutional Oversight: The Artificial Intelligence and Advanced Technology Council (AIATC) oversees AI strategy and regulation.

²³ <https://uaelegislation.gov.ae/en/policy/details/the-uae-charter-for-the-development-and-use-of-artificial-intelligence>

²⁴ <https://www.adiaic.gov.ae/en/home>

²⁵ <https://sdaia.gov.sa/>

²⁶ <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach>

²⁷ https://assets.publishing.service.gov.uk/media/6424208f3d885d000cdadddf/uk_ai_regulation_impact_assessment.pdf

национальных чемпионов, таких как Mistral AI, и в особом внимании к регулированию биометрических систем. Французская позиция во многом определяет жесткость подхода ЕС к базовым моделям в AI Act.

3.4 Американский континент

Закон Калифорнии CCPA (California Consumer Privacy Act, 2018) и его расширения CPRA (California Privacy Rights Act, 2020, вступил в силу в 2023) не регулируют ИИ или LLM напрямую. Это законы о конфиденциальности данных, которые де-факто становятся мощным инструментом регулирования бизнес-моделей, основанных на данных, включая LLM. Их влияние на LLM реализуется через контроль над "сырьем" (данные для обучения) и "выходом" (персональные выводы и решения).

Закон касается коммерческих организаций, которые ведут бизнес в Калифорнии и обрабатывают персональные данные более 50 тыс. потребителей, домохозяйств или устройств (порог CPRA) или имеют доход свыше 25 млн. долларов США, или получают более 50% дохода от продажи/обмена персональными данными. Практически любая крупная компания, разрабатывающая или развертывающая LLM, подпадает под эти критерии, если предоставляет услуги жителям Калифорнии.

Применительно к регулированию LLM представляет интерес положение о том, что промпт, введенный пользователем, является персональными данными, а профиль, который LLM может построить на основе диалога также сам по себе при определенных условиях можно отнести к защищаемым персональным данным.

За потребителями закреплены следующие права - право знать, право на доступ и право на удаление (Right to Know/Access/Delete) – это безусловно влияет на правила игры для LLM.

Компания должна раскрыть не только промпты пользователя, но и возможные выводы, сгенерированные моделью о пользователе (это технически сложно, так как выводы часто не хранятся явно, а являются частью весов модели). Если модель делает о пользователе ложный вывод, пользователь может потребовать его исправления. Как технически изменить внутреннее представление модели о конкретном человеке — также открытый вопрос.

Право на ограничение использования и раскрытия "чувствительных" данных (Right to Limit): CPRA вводит категорию "чувствительные персональные данные" (расовая/этническая принадлежность, здоровье, сексуальная ориентация, точное геоположение, содержимое частных сообщений). Для их обработки (включая "продажу") требуется согласие пользователя. Если пользователь делится с LLM информацией о своем диагнозе или религиозных взглядах — это обработка чувствительных данных.

Право на отказ от "продажи" и "обмена" данными: предоставляется право отказаться от "продажи" данных. CPRA расширил это право на "обмен" (sharing) данными с третьими сторонами в целях кросс-контекстной поведенческой рекламы, даже без денежной оплаты. Исходя из этого, передача промптов и результатов/выходных решений API-партнерам или использование диалогов для улучшения модели третьими сторонами может быть квалифицировано как "обмен" и потребовать механизма отказа.

Automated Decision-Making: CPRA вводит право потребителей получать информацию о логике автоматизированных решений и отказаться от "профилирования" в контексте доступа к финансовым и страховым услугам, жилью, трудоустройству, образованию. Так, если LLM используется для скрининга резюме, оценки кредитоспособности или вынесения судебных решений, компания обязана обеспечить прозрачность и возможность отказа. Однако объяснимость решений LLM остается сложной проблемой, которую трудно формализовать документально в жестких юридических терминах.

В итоге документ усложняет сбор и очистку обучающих данных из открытых источников (Интернет), так как они содержат персональные данные, подлежащие удалению по запросу. Повышенные требования к "чувствительным данным" делают рискованным использование медицинских, финансовых и других специализированных LLM без построения прозрачных систем для получения согласия и дальнейшего управления полученными данными.

ССРА/CPRA, не являясь законами об ИИ, становятся важным инструментом регулирования LLM через призму приватности. Они заставляют разработчиков и бизнес смещаться от модели "собрать все данные, спросить потом" к модели "минимальная необходимая коллекция, прозрачность и контроль пользователя с самого начала". Как уже отмечено, это создает значительные технологические и юридические вызовы, в области «машинного забывания» (machine unlearning), объяснимости (XAI) и управления согласием.

Следует также отметить, что создано Калифорнийское агентство по защите конфиденциальности (California Privacy Protection Agency, CPPA)²⁸ — первый в США специализированный регулятор в сфере приватности, который обладает полномочиями по принятию подзаконных актов и наложению штрафов. Ключевая идея состоит в том, что до создания CPPA в США не было отдельного государственного органа, чьей единственной задачей было бы регулирование защиты персональных данных на уровне штата или федерации.

Канада разрабатывает национальный закон об ИИ²⁹. Проект предполагает введение категоризации систем по уровню риска, запрет на определенные манипулятивные и дискриминационные практики, что напрямую касается дизайна и внедрения LLM на канадском рынке.

В Бразилии, где уже действует закон о защите данных LGPD (во многом аналогичный европейскому GDPR), ведётся работа над профильным Законопроектом об искусственном интеллекте. Анализ этого развивающегося акта важен для понимания того, как страны с законами о данных достраивают над ними специализированное регулирование ИИ, создавая дополнительный регуляторный слой для технологий ИИ.

Ниже приводятся ключевые положения и особенности Закона о защите персональных данных Бразилии (Lei Geral de Proteção de Dados Pessoais, LGPD, Закон № 13.709/2018)³⁰:

²⁸ <https://cppa.ca.gov/>

²⁹ <https://www.parl.ca/legisinfo/en/bill/44-1/c-27>

³⁰ <https://www.gov.br/esporte/pt-br/acao-a-informacao/lgpd>

Основные цели и сфера действия. Закон предполагает универсальное применение, он регулирует обработку персональных данных физических лиц, осуществляемую на территории Бразилии, независимо от местонахождения компании-оператора, а также обработку данных лиц, находящихся в Бразилии. Цель - гарантировать право на неприкосновенность частной жизни и защиту персональных данных, создать правовые рамки для ответственной обработки данных, добиться оптимального баланса между защитой прав граждан и инновациями в экономике.

Ключевые принципы обработки данных в LGPD:

- добросовестность и прозрачность;
- соответствие целям;
- адекватность и необходимость (минимизация сбора только до необходимых данных);
- свободный доступ;
- качество данных (точность и актуальность);
- безопасность;
- предупреждение возможного вреда;
- недискриминация;
- подотчетность.

LGPD устанавливает определенный ряд правовых оснований. Для законной обработки необходимо наличие как минимум одного из них:

- согласие субъекта данных (с определенными требованиями к форме и возможности простого отзыва);
- исполнение договора или преддоговорных процедур;
- выполнение правового или нормативного обязательства оператора;
- осуществление прав в судебных, административных или арбитражных процессах;
- защита жизни или физической безопасности субъекта данных или третьего лица;
- защита здоровья (в процедурах, осуществляемых медицинскими работниками);
- законные интересы оператора или третьей стороны (требуется баланс с правами субъекта данных);
- осуществление общественных интересов государственными органами.

Закон предоставляет физическим лицам широкие права, включая:

- подтверждение факта обработки (право узнать, обрабатываются ли его данные);
- доступ (право получить доступ к своим данным);
- исправление (право исправить неполные, неточные или устаревшие данные);
- анонимизация, блокирование или удаление (право потребовать удаления данных, обрабатываемых на основе согласия, или прекращения обработки);
- переносимость (право получить данные в структурированном формате для передачи другому поставщику услуг);

- информация о государственных и частных субъектах, с которыми оператор делился данными;
- информация о возможности не давать согласие и о последствиях отказа;
- отзыв согласия.

Контролер данных определяет цели и средства обработки персональных данных и несет основную ответственность за соблюдение LGPD. Обработчик данных обрабатывает данные от имени Контроллера по его указаниям. Назначение сотрудника по защите данных для большинства организаций предусмотрено в обязательном порядке, он выступает в качестве связующего звена между организацией, субъектами данных и Национальным управлением по защите данных (ANPD).

Передача данных за границу разрешена при соблюдении определенных условий; - в страны с уровнем защиты, адекватным LGPD (по решению ANPD); - при наличии гарантий, таких как стандартные договора, корпоративные глобальные правила или сертификаты; - с согласия субъекта данных.

Национальное управление по защите данных (Autoridade Nacional de Proteção de Dados, ANPD) — государственный орган, отвечающий за надзор, применение санкций, разработку нормативных актов и распространение информации о LGPD. Обладает административными и консультативными полномочиями.

ANPD может применять следующие санкции (после административного процесса):

- предупреждение с указанием срока для исправления;
- штрафы до 2% от оборота компании в Бразилии за последний финансовый год (исключая налоги), ограниченные 50 млн бразильских реалов (около 9 млн долларов США) за одно нарушение;
- публикация информации о нарушении;
- блокировка или удаление данных, связанных с нарушением;
- частичный или полный запрет на осуществление деятельности, связанной с обработкой данных.

LGPD во многом вдохновлен европейским Общим регламентом по защите данных (GDPR) и имеет схожие принципы и права. Однако есть и отличия, например, более широкий перечень правовых оснований (включая "законные интересы" и "защиту кредита") и особенности в структуре надзорного органа.

Применение ко всем секторам: в отличие от некоторых законов, LGPD применяется как к частному, так и к государственному сектору. Закон учитывает бразильскую правовую традицию, включая положения, связанные с публичным доступом к информации. Для обработки наиболее чувствительных категорий данных (информация о расе, религии, здоровье, биометрические данные и т.п.) применяются более строгие правила.

Таким образом LGPD установил в Бразилии современную, основанную на правах и принципах систему защиты данных, которая в т. ч. усиливает подотчетность (accountability) организаций. Компании, работающие в Бразилии или с данными бразильцев, должны построить программу

соответствия, уделяя особое внимание правовым основаниям, прозрачности и правам субъектов данных.

4. СТАНДАРТИЗАЦИЯ В СФЕРЕ ПРИМЕНЕНИЯ LLM

В условиях стремительной технической модернизации и внедрения больших языковых моделей (LLM) формирование единых, признанных на международном уровне стандартов и рамок для их оценки и ответственного управления становится критической необходимостью.

Международная организация по стандартизации (ISO) в сотрудничестве с Международной электротехнической комиссией (IEC) разработала комплекс взаимосвязанных стандартов, образующих целостную экосистему для управления искусственным интеллектом. Эти документы предоставляют организациям не разрозненный набор инструментов, а системную методологию, охватывающую весь жизненный цикл моделей ИИ — от концепции и разработки до внедрения, мониторинга и выводы из эксплуатации.

Центральным элементом этой архитектуры выступает стандарт ISO/IEC 42001:2023 «Системы управления искусственным интеллектом». Это первый в мире стандарт, предлагающий требования к созданию, внедрению, поддержанию и непрерывному улучшению системы менеджмента ИИ. Применительно к LLM, ISO/IEC 42001 задает организационный «каркас», в рамках которого компания может демонстрировать ответственный подход к использованию моделей. Он обязывает организацию формализовать политики, распределить роли и ответственность, а главное — интегрировать процессы управления уникальными рисками, присущими LLM, такие как генерация недостоверного контента, скрытые предубеждения в обучающих данных, вопросы интеллектуальной собственности и этические последствия применения.

Эффективное управление невозможно без общего языка и понимания базовых принципов, на которых построены технологии ИИ. Здесь фундаментальную роль играет стандарт ISO/IEC 22989:2022, устанавливающий единую терминологию и описывающий концепции и жизненный цикл систем ИИ. Для оценки LLM он является полезным, поскольку определяет ключевые понятия, такие как «генеративный ИИ», «обучение с учителем/без учителя» и другие, что обеспечивает однозначность трактовки в ходе аудита или оценки. Кроме того, описание этапов жизненного цикла ИИ формирует базовую структуру, к которой привязываются требования других стандартов.

С технологической точки зрения стандарт ISO/IEC 23053:2022, описывающий рамочную структуру для систем ИИ на основе машинного обучения, предоставляет концептуальную модель для анализа архитектуры LLM. Он помогает экспертам и оценщикам декомпозировать сложную систему на ключевые компоненты: данные, процессы обучения (включая обучение базовой модели и последующие этапы), алгоритмы оптимизации и итоговую модель. Это позволяет проводить более предметную оценку качества данных, воспроизводимости процесса обучения и технической документации.

Управление рисками, являющееся ядром ISO/IEC 42001, получает свое детальное развитие в руководстве ISO/IEC 23894:2023. Этот документ предлагает конкретные методологии для идентификации, анализа, обработки и мониторинга рисков, специфичных для ИИ. В контексте LLM он помогает понять, как системно подходить к оценке вероятности и воздействия таких

событий, как утечка конфиденциальных данных из промптов, нарушение приватности при дообучении на корпоративных данных, манипулятивное поведение модели или нанесение репутационного ущерба из-за сгенерированного контента. Процесс, описанный в этом стандарте, напрямую поддерживает выполнение требований по управлению рисками, предъявляемых ISO/IEC 42001.

В целом система управления ИИ должна быть интегрирована с существующими в организации системами менеджмента. Поэтому стандарты ISO/IEC 27001 (информационная безопасность) и ISO 9001 (менеджмент качества) выступают критически важными комплементарными элементами. ISO/IEC 27001 обеспечивает защиту на всех этапах работы с LLM: будь то исходные обучающие данные (датасеты), промпты пользователей, содержащие коммерческую тайну, или промежуточные результаты вычислений. ISO 9001 позволяет встроить процессы управления LLM в общий цикл непрерывного улучшения качества продукции и услуг, обеспечивая их предсказуемость и соответствие целям бизнеса.

Современная система стандартов ISO/IEC для ИИ представляет собой целостную, многоуровневую экосистему. Ее применение для оценки LLM позволяет перейти от точечных технических тестов к комплексному организационному аудиту. Она предоставляет заказчикам, регуляторам и обществу прозрачный механизм для верификации того, что сложные и потенциально рискованные технологии, такие как большие языковые модели, разрабатываются и используются в рамках ответственной системы управления, нацеленной на минимизацию вреда и максимизацию общественной пользы.

На техническом уровне в части обеспечения прозрачности автономных систем к проблеме использования LLM может быть применен также стандарт **Института инженеров электротехники и электроники (Institute of Electrical and Electronics Engineers - IEEE) 7001-2021 - IEEE Standard for Transparency of Autonomous Systems**³¹. Документ предлагает детальные, инженерно-ориентированные требования к объяснимости, которые могут быть прямо использованы регуляторами для оценки LLM. Эти требования в т. ч. могут быть заложены в контракты для государственных закупок.

Данный стандарт широко применим к автоматизированным системам управления, системам медицинской диагностики, рекомендательным системам и чат-ботам. Особый интерес для данного стандарта представляют автономные системы, которые потенциально могут причинить вред. В сферу действия входят системы, критически важные для безопасности. Данный стандарт рассматривает системы, способные непосредственно причинить физический, психологический, социальный, экономический или экологический вред, а также ущерб репутации.

³¹ IEEE Standard for Transparency of Autonomous Systems," in *IEEE Std 7001-2021*, vol., no., pp.1-54, 4 March 2022, doi: 10.1109/IEEESTD.2022.9726144; <https://ieeexplore.ieee.org/document/9726144>

5. НАУЧНЫЕ РАБОТЫ ПО ПРОБЛЕМЕ РЕГУЛИРОВАНИЯ LLM

Изучение проблемы регулирования больших языковых моделей (LLM) требует обращения к исследованиям и научным работам, которые анализируют ядро проблемы и предлагают концептуальные решения.

Ниже приводятся некоторые тезисы, которые являются центральными и обобщают наиболее важные акценты в современных научных работах.

а) Старые механизмы регулирования не совместимы с новыми всеобъемлющими технологиями. В ряде исследований выдвигается тезис о том, что традиционные правовые режимы (основанные на территориальности, предсказуемости действий и причинно-следственной связи) структурно несовместимы с глобальными, самообучающимися и непрозрачными системами ИИ, включая LLM. Такой взгляд задаёт теоретическую рамку, объясняющую, почему простое "наложение" старых законов на новые технологии обречено на провал.

б) Регулирование не на уровне конкретных приложений, а на уровне "модельного слоя". Отмечается необходимость лицензирования разработчиков базовых моделей, стандартов аудита и раскрытия информации. Регулировать нужно не тысячи конкретных приложений, а немногочисленных разработчиков фундаментальных моделей. Требовать от них лицензирования, аудита безопасности и раскрытия информации о данных и архитектуре. Речь в т. ч. идет о создании специальных регуляторных "шлюзов" для контроля доступа к мощным моделям.

Предлагается также сместить фокус с регулирования самих алгоритмов на регулирование критической инфраструктуры (облачные платформы, вычислительные кластеры, API-маркетплейсы и т. п.), что позволит осуществлять более эффективный и централизованный контроль над развитием и распространением LLM, избегая необходимости "заглядывать внутрь черного ящика". Так, например государства должны концентрировать контроль на критически важных узлах: облачных платформах, вычислительных кластерах и маркетплейсах API.

в) Акцент на защите интеллектуальной собственности, вопросах ответственности и правовой рамке в области данных. Проблемы, связанные с обучением моделей на данных, охраняемых авторским правом, а также неопределенность статуса сгенерированного контента все чаще поднимаются в дискурсе вокруг LLM. В ряде исследований проводится анализ может ли доктрина "добросовестного использования" (fair use) применена создателям LLM. Исследуются вопросы гражданско-правовой ответственности за вред, причиненный LLM, а также возможности создания систем обязательного страхования для разработчиков высокорисковых LLM. В рамках таких исследований, в частности, высказываются идеи о том, что разработчики высокорисковых LLM должны нести финансовую ответственность за причиненный вред, предлагаются механизмы обязательного страхования, аналогичные тем, что действуют в ядерной или авиационной отрасли – все это поможет правильно распределять риски.

г) Предлагается более широко применять практику внедрения государственных "санбоксов" (regulatory sandboxes), т. е. создавать контролируемые юридические зоны для тестирования LLM перед выходом на рынок. Это позволит оценивать риски в реальных условиях, не блокируя инновации.

д) Предложения по усилению роли технических стандартов, особенно при обеспечении требований прозрачности. Законы должны опираться на конкретные технические стандарты (ISO/IEC, IEEE) в области аудита данных, смещенности, объяснимости и др.

е) В качестве важного направления предлагается прямой государственный контроль над критическими вычислениями (compute governance), а также мониторинг за продажей/поставками и квотирование продаж специализированных чипов (GPU). А кроме того, введение строгой подконтрольности в вопросах доступа к суперкомпьютерам и к большим вычислительным ресурсам/центрам. Таким образом контроль над LLM можно вести и за счет контроля за вычислениями большого масштаба, которые требует сегодня значительных мощностей, концентрации технологических инноваций и талантов.

Пожалуй, наиболее важным выводом на основе анализа научных статей является наметившийся тренд, связанный с переходом от регулирования поведения моделей к контролю над условиями их создания, условиями доступа и эксплуатации MML.

6. НЕКОТОРЫЕ ПРАКТИКИ И РЕЖИМЫ ДОСТУПА К LLM

- практика регистрации, когда доступ к LLM организован через многоуровневую систему, обеспечивающую контроль, монетизацию (по решению провайдера) и безопасность;
- Freemium-модель с обязательной регистрацией (наиболее распространена); Freemium-модель основана на бизнес-стратегии, предусматривающей предоставление базовых услуг бесплатно, но за расширенные функции взимается плата, таким образом часть пользовательской базы превращается в платных клиентов;
- корпоративные контракты в вендором, юридическое соглашение о предоставлении услуг, гарантии уровня обслуживания (SLA), расширенные настройки безопасности, приватное развёртывание, другие требования заказчика/пользователя;
- геоблокировка и ограничение доступа к LLM в странах, находящихся под санкциями или где нет юридической ясности; реализуется через блокировку IP-адресов или требований к определенным способам оплаты; соответственно доступ к ряду LLM может быть обусловлен правилами трансграничных онлайн-платежей;
- налоговые требования (в т. ч. используется как инструмент регулирования); страны стремятся взимать налог с цифровых услуг, потребляемых на их территории, даже если поставщик находится за рубежом

Два основных подхода для взимания НДС/налога с продаж с международных транзакций за LLM:

- Механизм "обратного начисления" (Reverse Charge Mechanism): обязанность по уплате НДС перекладывается с иностранного поставщика на местного корпоративного потребителя (B2B).
- Регистрация иностранных поставщиков и уплата налога у источника (Direct Registration & OSS): иностранный поставщик цифровых услуг (LLM)

обязан зарегистрироваться в качестве налогоплательщика в стране потребления и уплачивать НДС напрямую в её бюджет.

Некоторые страны (например, Турция, Саудовская Аравия) обязывают платежные системы и банки удерживать налог при переводе средств за цифровые услуги за рубеж.

В итоге можно отметить, что доступ к LLM может регулироваться не только условиями провайдеров, но и санкционным, валютным и налоговым правом. Налоговые органы активно адаптируют инструменты для цифровой экономики, стремясь извлекать фискальную выгоду из потребления зарубежных ИИ-сервисов и ставя их в равные условия с местными компаниями.

7. ОБЩИЕ ВЫВОДЫ

1. В мире наблюдается фундаментальная проблема совместимости различных национальных режимов регулирования ИИ и, в частности, регулирования LLM.

Условно и в весьма обобщенном виде можно выделить следующие подходы к регулированию:

- инновационно-инвестиционная модель (ОАЭ, Саудовская Аравия);
- риск-ориентированная (ЕС, Канада);
- принцип-ориентированная (Великобритания, Япония);
- суверенно-изоляционная (Китай, Россия).

2. Представленная грубая классификация подходов, однако, не учитывает все нюансы региональных и страновых законодательств и конкретные правоприменительные практики в странах. Это безусловно требует более детального изучения всей мозаики регуляторных механизмов и намечающихся трендов в тех юрисдикциях, которые уже активно приступили к законодательной работе в обозначенном направлении.

3. Очевидно, что возможности и функции LLM, широкий спектр их применения в различных сферах и обширные способы взаимодействия компаний и потребителей (граждан) с такими моделями показывают, что в регулировании LLM мы неизбежно заходим в следующие регуляторные области:

- весь комплекс вопросов цифрового суверенитета и информационной безопасности, особенно это касается защиты персональных данных и приватности;
- защита данных на всех этапах передачи, обработки и хранения (оборот данных), включая порядок использования облачных хранилищ и требования к локализации серверов;
- регулирование в области цифровых услуг;
- защита интеллектуальной собственности и патентная политика;
- налоговая система и соответствующие нормативные требования;
- регулирование информационной среды и работы СМИ;
- регулирование социальных сетей и цифровых платформ, в т. ч. в части рекомендательных алгоритмов;

- развитие стандартизации и оценки соответствия, сертификация моделей;
- защита традиционных ценностей, родного языка, самобытной культуры;
- соответствие цифрового продукта принятым этическим нормам, моральным принципам и стандартам общественного поведения;
- гуманитарные аспекты, связанные с воздействием LLM на человека и общество, научное исследование долгосрочных эффектов и последствий применения технологии.

4. Исходя из взаимопроникновения областей регулирования, работа по созданию законодательной базы и нормативного контура для LLM требует, помимо прочего, гармонизации с уже имеющимися нормативными положениями и тщательного анализа конкретного предмета регулирования в рамках вышеперечисленных областей законотворчества (пункт 3). Разработка законов предполагает и анализ технологических процессов и возможностей LLM различного назначения в различной среде, включая практики их отраслевого применения (особенно выверенно необходимо действовать в медицине, образовании, судебной системе, госуправлении).

5. Необходимо иметь в виду, что на национальном уровне активно создаются различные отраслевые и прикладные/профильные (специализированные под различные задачи) системы LLM и всевозможные сервисы на основе генеративного ИИ (в масштабах отрасли, корпоративные модели в компаниях и организациях, бизнес-сервисы в ретейле и т. п.). В этой связи режимы ограничения зарубежных/транснациональных моделей при недостаточной оценке последствий могут негативно повлиять на развитие ИИ в нашей стране и замедлить программы цифровой экономики, особенно в реальном секторе производства, где создаются как собственные модели LLM, так и на основе общеизвестных мировых платформ. Это предполагает тщательную проработку и всестороннюю выверенность законодательных инициатив.

Данный обзор предлагает предварительный общий анализ некоторых подходов и практик регулирования LLM за рубежом, который может быть расширен и углублен в рамках отдельных направлений и законотворческих задач.