Информационно-аналитическая справка по отчету

«Роль правительства как поставщика данных для ИИ»

Национальный центр развития искусственного интеллекта при Правительстве Российской Федерации

Наименование отчета: «Роль правительства как поставщика данных для ИИ» («The Role of Government as a Provider of Data for Artificial Intelligence»)

Разработчик отчета: Global Partnership on Artificial Intelligence, США

Дата выпуска отчета: май, 2024

Объем отчета: 63 стр.

Тэги: #Мировые рейтинги, индексы ИИ #2024 #США

Ссылка на скачивание отчета из оригинального источника: https://africa.ai4d.ai/wp-content/uploads/2024/01/DG08-The-Role-of-Government-as-a-Provider-of-Data-for-Artificial-Intelligence-Interim-Report.pdf

Ссылка на скачивание отчета в Базе данных: https://ai.gov.ru/knowledgebase/investitsionnaya-aktivnost/2024_roly_pravitelystva_kak_postavschika_dannyh_dlya_ii_the_role_of_government_as_a_provider_of_data_for_artificial_intelligence_global_partnership_on_artificial_intelligence/

Справочно: Глобальное партнерство по искусственному интеллекту (GPAI) объединяет членов ОЭСР для продвижения повестки по внедрению ориентированного на человека, безопасного, надежного и заслуживающего доверия искусственного интеллекта ИИ.

Основные разделы отчета:

- Концепция доклада
- Государство как поставщик данных
- Анализ практик стран
- Ключевые факторы обмена государственными данными
- Устранение препятствий для доступа к государственным данным
- Рекомендуемые принципы обмена данными

Ключевые тезисы отчёта:

- Правительства **имеют ограниченные возможности для оценки и проверки данных, а также контроля в вопросах использования данных, которыми они делятся с разработчиками ИИ**. Это особенно актуально для стран с низким и средним уровнем дохода.

- Правительственные **порталы открытых данных остаются главными ресурсами, через которые можно организовать прозрачный обмен правительственными данными для разработки ИИ**. Данные на порталах должны актуализироваться своевременно. Не менее важно и то, что такие порталы это не просто сайты для хранения данных, а ресурс для установления норм и правил регулирования обмена данными.

- **Расширение прав и возможностей субъектов данных в соглашениях об обмене данными между правительствами и разработчиками ИИ является первостепенной задачей**. Это сопряжено с ограничениями, которые невозможно избежать, но которые необходимо соблюдать.

- Важнейший шаг в построении эффективной модели обмена и применения государственных данных является **повышение уровня доверия граждан к ИИ, их осведомленности о механизмах работы ИИ.**

1. Основное содержание разделов Отчета

Концепция доклада

- Правительства являются важными агрегаторами данных. **Только они располагают данными, которые охватывают всю страну**. Эти данные могут быть важной основой для разработки инструментов ИИ, которые улучшат жизнь граждан.

- В вопросе распределения данных правительство может выступать в роли поставщика данных и поставщика доступа к ним (См. Рис.1).

Рис. 1. Роли правительства как поставщика данных

Каждый из способов связан с определенным набором правил или параметров и позволяет подходить более гибко к вопросам регулирования использования данных в стране**.** Ключевым вопросом регулирования обмена данных является **концепция ответственной открытости**. Данные правительства должны предоставляться разработчикам таким образом, который бы не подвергал риску граждан страны – субъектов данных и не тормозил развитие ИИ-проектов.

- Большие ИИ-компании, используя свое положение, **пренебрегают мнением субъектов данных**. Например, компания по распознаванию лиц CloudWalk заключила сделку с правительством Зимбабве по разработке биометрической системы распознавания.

Правительство страны не озаботилось подготовкой должного нормативного регулирования, **что позволило компании присвоить себе базу данных лиц граждан**. Законы Зимбабве, регулирующие персональные данные и трансграничную передачу данных, отсутствовали, Правительство не смогло обеспечить защиту прав граждан.

- Правительства не стремятся придерживаться принципов открытости в вопросах использования данных ИИ-компаниями. **Проблема отсутствия общедоступной информации по вопросам обмена данными прослеживается везде в мире**. Например, Royal NHS Free Trust поделился данными пациентов с Google Health для использования в приложении. Эта информация стала общедоступной только из-за расследования после того, как были поданы жалобы на то, что частные данные пациентов передавались без надлежащих разрешений. Основополагающая документация по процессу обмена данными была закрыта.

Государство как поставщик данных

Управление данными касается **не только защиты, но и доступа, безопасности и доверия**. Существует четыре основных способа, с помощью которых государственные данные могут быть предоставлены разработчикам ИИ.

1. Открытые данные — наиболее известный способ предоставления правительственных данных. Открытые данные содействуют равному доступу к правительственным данными, как правило, бесплатно. Существуют принципы, регулирующие открытые данные, такие как недискриминационный доступ и возможность повторного использования данных. Поскольку открытые данные означают неограниченный публичный доступ, личные данные не могут передаваться через этот механизм.

2. Институты данных — это организации, которые направляют использование данных в интересах общественности. Они необязательно должны быть государственными. В рамках этой модели правительства могут разрабатывать и поддерживать стандарты и инфраструктуру для обмена и работы с данными, а непосредственный обмен данными осуществляется в том числе субъектами данных. Для эффективной работы институтов данных требуется признание прав субъектов данных, таких как право на переносимость данных. Переносимость данных позволит индивидуально обмениваться персональными данными с разработчиками ИИ при содействии правительств, но она не позволит осуществлять массовый доступ к данным, и для массового обмена данными необходимы альтернативные модели, такие как государственно-частное партнёрство.

3. **Государственно-частное партнёрство** для гармонизации услуг, предлагаемых субъектам данных. Например, вместо того чтобы повторно запрашивать у пользователей информацию, которую они уже предоставили государственной системе, данные могут быть повторно использованы (на основе согласия пользователей) во время заявок на кредит. Данные хранятся в цифровой инфраструктуре, которая связывает данные государственных учреждений, муниципалитетов и регистров. В таком случае цифровая инфраструктура поддерживается частным образом.

4. Контрактная система обмена данными, базирующаяся на договорах. Может охватывать различные области по работе с данными: описание продукта или услуг, которые должны быть

исполнены на основе данных; регламент сбора, доступа и контроля данных; характер, цель и продолжительность обработки данных; типы данных и категории субъектов данных; обязательства и права исполнителя и правительства; местоположение данных и описание системы доступа; хранение данных и т.д. При такой системе правительство делится конкретным набором данных под конкретную задачу для конкретного исполнителя.

- Все четыре модели обмена данными имеют свои преимущества и недостатки. Открытые данные вызывают опасения по поводу свободного доступа к данным. Это можно решить, установив законодательные рамки на использование данных. Модель государственно-частного партнерства может привести к **неравному доступу к данным через предоставление конкурентного преимущества отдельным компаниям**. Договорные соглашения между правительствами и частным сектором также сопряжены с рисками, которые можно смягчить, если соглашения об обмене данными будут проработанными.

Анализ практик стран

В рамках данного исследования был проведен анализ четырех кейсов применения ИИ в государственном секторе. Выводы по кейсам были обобщены в несколько рекомендаций:

- Правительства делятся своими данными с ИИ-разработчиками для решения социальных проблем. **ИИ-инструменты повышают эффективность госсектора, но есть риск подорвать доверие населения, если правительства делятся конфиденциальными данными с ИИ-разработчиками вне принятых и прозрачных норм и правил**. Субъекты данных должны быть осведомлены о том, кем и для каких целей применяются данные о них.

- **Развитие ИИ выигрывает от объединения данных государственного сектора с данными частного сектора**. Эффективный обмен данными требует надзора, правил и аудита. Интеграция разнообразных наборов данных создает полезные базы данных для обучения ИИ. Но и в этом случае проблема конфиденциальности данных должна стоять на первом месте.

- Применение ИИ в госсекторе улучшает автоматизацию и позволяет упростить и ускорить принятие решений, но может привести **к ошибкам, дискриминации и предвзятости ИИ**. А значит **полная опора на автоматизированные решения невозможна** и это надо учитывать. Данные должны собираться, распространяться и использоваться ответственно и этично. ИИ может стимулировать социально-экономические преобразования, но правительства, как главные распорядители данных, должны гарантировать, что это делается без риска для людей.

- Авторы исследования заявляют о необходимости сильного государственного надзора для обеспечения соблюдения законов и постановлений, а также предотвращения обхода обязательств частным сектором. Без механизмов надзора **частный сектор будет стремиться игнорировать правила и постановления, так как они накладывают на него дополнительные издержки**.

Ключевые факторы обмена государственными данными

- Перед тем как передавать данные разработчикам ИИ, государство должно создать три сферы: нормативно-правовую, которая будет обеспечивать равный доступ к данным сторонним

разработчикам ИИ; политическую, которая будет облегчать обмен данными внутри и за пределами правительства, и инфраструктурную для обеспечения доступа к государственным данным.

Нормативно-правовая сфера

- ✓ **Доступ к государственным данным важен для ИИ-разработчиков, так как это даёт конкурентное преимущество. Крупные разработчики ИИ могут использовать свой доступ к государственным данным, становясь эксклюзивными подрядчиками для монополизации рынка. Поэтому правительствам следует таким образом регулировать оборот данных, чтобы предотвратить монополизм и зависимость от одного разработчика.**
- ✓ **Государства вводят ограничения на трансграничную передачу данных, не заботясь о целесообразности запретов. Эффективное регулирование оборота данных должно разграничивать типы данных (например, персональные и не персональные, конфиденциальные и не конфиденциальные) для соблюдения баланса между защитой данных и развитием ИИ. Например, инициатива по трансграничной передаче данных, которая была принята в рамках АТЭС соблюдает этот баланс.**
- ✓ **Государственные данные могут содержать информацию, которая защищена правом интеллектуальной собственности. Например, международное соглашение по торговым аспектам прав интеллектуальной собственности защищает такую информацию. Важно установить четкие правила по вопросу авторских прав на проекты ИИ, созданные на основе таких данных. Это может быть решено через систему лицензий, в том числе и лицензий стандарта Creative Commons.**
- ✓ **Для эффективного раскрытия данных необходимы механизмы, которые обеспечивают не только прозрачность, но и диалог с гражданами. Законы о доступе к информации играют в этом ключевую роль, позволяя людям получать информацию о том, какие данные хранятся, как они защищаются, используются и кому передаются.**

- Политика играет ключевую роль в продвижении обмена данными, особенно в государственном секторе, **создавая условия для работы нормативной и инфраструктурной сфер**. Политика должна учитывать приоритеты страны и поддерживать развитие технологий, соответствующих этим приоритетам. Примером является Рамочная политика данных Африканского союза. **Совместимость стандартов создания, хранения и обмена данными, а также их прозрачность – это важные цели государственной политики в области данных**.

- **Обеспечение доступности, конфиденциальности и безопасности данных — это задачи инфраструктурной сферы**. Инфраструктура данных включает в себя процессы хранения, обработки, управления, доступа и анализа данных, а также требует широкой цифровой инфраструктуры. Разные типы данных требуют разных подходов к хранению. Локальные и облачные серверы играют ключевую роль, при этом вопросы безопасности и суверенитета данных должны оставаться в приоритете. Доступ к данным через API способствует большей применимости данных. Инфраструктура для обмена данными должна учитывать цифровые

возможности пользователей, будь то государство, частный сектор или гражданское общество. Инвестиции в инфраструктуру данных должны учитывать контекст использования.

Устранение препятствий для доступа к государственным данным

Уникальные особенности стран порождают различные проблемы в процессах обмена государственными данными. Однако существует пять единых для всех барьеров:

1. **использование персональных данных для обучения систем ИИ;**
2. **нормативно-правовая система;**
3. **система оценки стоимости;**
4. **технические возможности\ограничения;**
5. **отношение населения к ИИ.**

Рекомендации ниже предлагают способы смягчения этих барьеров.

- Использование больших наборов государственных данных для разработки ИИ вызывает опасения в области конфиденциальности, так как **они почти всегда содержат конфиденциальную информацию.** Неправильное обращение с такими данными может привести к нарушению законов о конфиденциальности данных. Технологии для защиты конфиденциальности: **анонимизация, дифференциальная конфиденциальность и синтетические данные, позволяют безопасно использовать данные для работы ИИ.** Анонимизация удаляет личную информацию из наборов данных. Синтетические данные, созданные с помощью алгоритмов, имитируют реальные данные, сохраняя все зависимости внутри данных, но сохраняют конфиденциальность. **Однако использование синтетических данных требует осторожности.** Европейская комиссия провела анализ возможных рисков и последствий использования синтетических данных (См. Таб.1).

Таблица 1. Негативные и позитивные факторы применения синтетических данных

Негативные предполагаемые последствия	Позитивные предполагаемые последствия
1. Один из способов улучшения точности ответов ИИ на основе синтетических данных — сравнение синтетических данных с оригинальными. **Однако для этого требуется доступ к оригинальным данным.** 2. Синтетические данные имитируют реальные данные, но не являются их копией. Некоторые методы генерации синтетических данных могут не учитывать отдельные особенности, которые присутствуют в оригинальных данных. Эти особенности могут оказаться важными для установления зависимостей. 3. Качество синтетических данных тесно	**1.** Синтетические данные по умолчанию конфиденциальные, а значит **проблем с нарушением конфиденциальности не будет.** **2.** Синтетические данные могут способствовать снижению предвзятости ИИ, **используя сбалансированные синтетические наборы данных для обучения.**

связано с качеством исходных данных и моделью генерации данных. Синтетические данные могут отражать предвзятость исходных данных. Манипулирование реальными наборами данных для создания сбалансированных синтетических наборов данных может привести к получению неточных данных.	

- **Правовые нормы не успевают модернизироваться в соответствии со скоростью появления новых технологий.** Для соблюдения баланса между развитием ИИ и поддержкой его безопасности для общества, правительства создают регуляторные песочницы. **Эти песочницы позволяют разрабатывать, тестировать и проверять модели ИИ до их выхода на рынок**. Работа песочниц требует участия различных акторов для учета технических, социальных и экономических особенностей новых технологий ИИ.

- **Открытые данные стран могут быть использованы в коммерческих целях.** А значит ответственное использование данных государственного сектора для развития ИИ требует **разработанных финансовых моделей**. На стоимость данных влияют их характер, структура и объем. Правительствам следует применять правила назначения стоимости на данные и распределения прибыли с учетом общественной полезности и справедливости. Важно заинтересовать разработчиков ИИ в выгодности применения данных.

- Государственный сектор может способствовать развитию ИИ **только при наличии необходимых человеческих ресурсов.** В условиях быстрого прогресса в области ИИ государственные служащие должны обладать навыками работы с ИИ и понимания особенностей его функционирования.

- Разработка и внедрение управления данными для ИИ должны следовать этическим принципам. При этом **этические нормы могут варьироваться в разных культурных контекстах**. Восприятие одной и той же технологии из сферы ИИ может быть разным в зависимости от страны. Важным аспектом является влияние политической культуры на практику обмена данными. **Политическая культура закрытости и секретности не способствует развитию практик обмена данными.**

Рекомендуемые принципы обмена данными

- **Общественная польза.** ИИ, использующий данные правительства, должен приносить пользу обществу и способствовать устойчивому развитию.

- **Подотчетность.** Использование данных должно быть полностью контролируемым и соответствовать законодательству.

- **Участие субъектов данных.** Субъекты данных должны давать согласие на передачу данных, кроме специальных случаев. Права субъектов должны быть защищены.

- **Равенство данных и справедливость**. Правительства должны обеспечивать равный доступ к данным через справедливые модели обмена и инфраструктуру.

- **Прозрачность**. Правительства должны быть прозрачны в отношении целей и методов обмена данными, привлекая общественность к обсуждению.

- **Безопасность.** ИИ должен быть технически безопасным и защищать данные от нарушений.

- **Минимизация данных**. Передача данных должна быть ограничена необходимым объемом для достижения цели.

- **Защита прав человека и конфиденциальности.** ИИ должен защищать права человека, исключать предвзятость и иметь независимый контроль.

Применимость отчета в Российской Федерации:

В результате анализа отчета, можно дать следующие рекомендации по развитию управления данными для ИИ:

- Государству следует вести разработку правили инфраструктуры по работе с данными одновременно. Также важным является повышение компетенций государственных служащих к работе с данными, обмену ими и открытости к новым практикам.

- Государственные данные следует рассматривать как продукт, требующий соответствующего обращения. Для этого необходимо привлекать разработчиков ИИ, устанавливать справедливые цены на приобретение ими данных, постоянно улучшать и развивать это сотрудничество. Важным шагом также является создание порталов открытых данных.

- Государственные данные следует рассматривать как продукт, требующий соответствующего обращения. Для этого необходимо привлекать разработчиков ИИ, устанавливать справедливые цены на приобретение ими данных, постоянно улучшать и развивать это сотрудничество. Важным шагом также является создание порталов открытых данных.