Информационно-аналитическая справка по отчету

«Отчет о состоянии ИИ в 2023 году»

Национальный центр развития искусственного интеллекта при Правительстве Российской Федерации

Наименование отчета: «Отчет о состоянии ИИ в 2023 году» («State of AI Report 2023»)

Разработчик отчета: Air Street Capital, США

Дата выпуска отчета: октябрь, 2023

Значимость отчета: 5 из 5

Объем отчета: 163 стр.

Тэги: #Мировые рейтинги, индексы ИИ #2023 #США

Ссылка на скачивание отчета из оригинального источника: https://www.stateof.ai/2023-report-launch

Ссылка на скачивание отчета в Базе данных: https://ai.gov.ru/wiki/komponenty/infrastruktura-ii/2023_otchet_o_sostoyanii_ii_state_of_ai_report_air_street_capital/

Справочно: Венчурный фонд «Air Street Capital» инвестирует в компании в области искусственного интеллекта (ИИ). Миссия фонда – создавать устойчивые компании, оказывающие долгосрочное влияние на рынки.

Основные разделы отчета:

- Введение
- Исследования
- Промышленность
- Политика
- Безопасность
- Прогнозы

10 прогнозов от авторов доклада на следующие 12 месяцев:
1. Кинопроизводство уровня Голливуда будет использовать генеративный искусственный интеллект для визуальных эффектов.
2. Медиа-компания, использующая генеративный ИИ, подвергнется расследованию за злоупотребление новыми технологиями во время выборов в США в 2024 году.
3. Самообучающиеся ИИ-модели превзойдут сегодняшние SOTA-модели (перспективные ИИ-технологии), особенно в сложной инфраструктурной и научной среде.
4. Рынки IPO в сфере технологий вступят в ИИ-среду, и произойдет, по крайней мере, один крупный листинг компании, ориентированной на искусственный интеллект (например, Databricks).
5. Глобальное увлечение масштабированием генеративного ИИ приведет к тому, что одна из компаний потратит >1 млрд долл. США на обучение одной крупномасштабной модели.
6. Федеральная торговая комиссия США или Управление по конкуренции и рынкам Великобритании займется расследованием сделки Microsoft/OpenAI о монополизации рынка.
7. Мы видим лишь небольшой прогресс в области глобального управления искусственным интеллектом, выходящего за рамки добровольных обязательств.
8. Финансовые учреждения запустят долговые фонды для полупроводников, чтобы финансировать проведение вычислений для развития ИИ вместо венчурных капиталистов.
9. Песня, сгенерированная искусственным интеллектом, попадет в топ-10 Billboard Hot 100 или Spotify Top Hits 2024.
10. Поскольку рабочая нагрузка и затраты, необходимые для вывода данных, значительно возрастают, крупная компания по ИИ (например, OpenAI) приобретет компанию по производству чипов для ИИ, ориентированную на решение задач вывода.

Ключевые тезисы отчета:

Исследования

- На примере GPT-4 виден существенный **рост популярности частных и закрытых (проприетарных) ИИ-моделей**, принадлежащих компаниям, обгоняющих по популярности открытые альтернативы. Этот же пример демонстрирует **высокую эффективность обучения с подкреплением на основе обратной связи от пользователей.**

- Увеличивается количество попыток повторить или превзойти качество больших языковых моделей (LLM) с помощью **небольших альтернатив**, применяемых **в отдельных отраслях или компаниях**, а также повышения качества датасетов.

- По некоторым оценкам, уже в 2025 году **LLM могут столкнуться с дефицитом данных**, в связи с чем могут возникнуть проблемы с развитием и масштабированием ИИ. Перспективным, но малоизученным остается **влияние синтетических данных на качество моделей**. Еще одним важным направлением является **извлечение новых данных из видео и введение в оборот закрытых данных компаний**. При этом все более подробно изучаются **характеристики контекста**, необходимые для извлечения наибольшей пользы из ИИ-модели.

- Большие языковые модели станут инструментом, обеспечивающим большой **прорыв в естественных науках, и в первую очередь в молекулярной биологии и создании новых белков**. В связи с этим, наибольший прирост числа научных публикаций по ИИ приходится на медицину (около 80 000).

Промышленность

- **NVIDIA достигла триллионной капитализации** за счет спроса на графические процессоры для ИИ в государственном секторе, стартапах, технологических компаниях и исследовательских организациях. Более того, **их чипы используются в исследованиях в сфере ИИ в 19 раз чаще**, чем все альтернативы вместе взятые.

- **Экспортные ограничения затрудняют продажи чипов на глобальном рынке** (в первую очередь в Китае), но ключевые вендоры создают специализированную продукцию для обхода таких ограничений.

- Приложения на основе **генеративного ИИ привлекли рекордные 18 млрд долл. США** венчурных и корпоративных инвестиций в 2023 году. **На компании, работающие с генеративным ИИ, пришлось на 33% больше начального финансирования и на 130% инвестиций в раунде A, чем на все стартапы в 2023 году**.

Политика

- 2023 год ознаменовался продолжающимся формированием **двух технологических полюсов в лице США и Китая.** Пока инициатива на стороне США, мобилизующих своих партнеров для

производства полупроводников, однако Китай продолжает наращивать объем собственных ИИ-разработок и исследований.

- Ограничив экспорт чипов в Китай, **США рискуют ослабить рынок для своих собственных производителей**.

Безопасность

- Дискуссия об **экзистенциальном риске от ИИ впервые вышла на первый план** в этом году. Это особенно важно, учитывая, что даже **высокопроизводительные модели легко поддаются «джейлбрейку»** (определенный вводный запрос для ИИ-модели, который позволяет обойти внутренние ограничения, установленные разработчиками для безопасности). Чтобы устранить эти проблемы, исследователи изучают **альтернативные варианты**, например, **предварительное обучение ИИ с учетом новых запросов человека, «вознаграждая» ИИ-модель за правильное «поведение».**

Основные тезисы подобнее:

1. **Исследования**

- **2023 год был годом больших языковых моделей**: GPT-4 от OpenAI стал мировой сенсацией, превзойдя все остальные LLM – как по классическим тестам для ИИ, так и по экзаменам, предназначенным для людей. OpenAI сообщили, что, хотя **GPT-4 по-прежнему страдает “галлюцинациями” (ошибочные действия ИИ), он генерирует правильные ответы на 40% чаще, чем предыдущие модели GPT.**

- **Обучение с подкреплением на основе обратной связи пользователей играет центральную роль в успехе современных LLM**, особенно тех, которые предназначены для прямого общения с людьми. К ним относятся Claude от Anthropic, Bard от Google, LLaMa-2-chat от Meta и, конечно, ChatGPT от OpenAI.

- Сегодня мы являемся свидетелями **отхода компаний от раскрытия информации о своих разработках на фоне проблем с безопасностью и конкуренцией**. OpenAI опубликовала лишь короткий технический отчет для GPT-4, как и Google – для PaLM2, при этом Anthropic не раскрыла вообще никакой информации о Claude. Единственным исключением может стать Meta и ее модель LLaMa, обученная исключительно на общедоступных наборах данных. Это оставляет возможность для развития иных моделей с открытым исходным кодом.

- **Длина контекста для обучения моделей становится все более важной темой исследований.** Исследователи обнаружили, что **производительность моделей была выше**, когда релевантная для задачи **информация появлялась в начале или в конце входных данных** с резким падением эффективности в середине. Они также обнаружили, что производительность модели снижалась по мере увеличения длины входных данных.

- Исследователи Microsoft показали, что обучение **небольших языковых моделей** с использованием специализированных и тщательно отобранных наборов данных даёт возможность **конкурировать с моделями в 50 раз больше по объему.** Они также обнаружили, что их технические аспекты **поддаются лучшей интерпретации**, решая проблему «черного ящика».

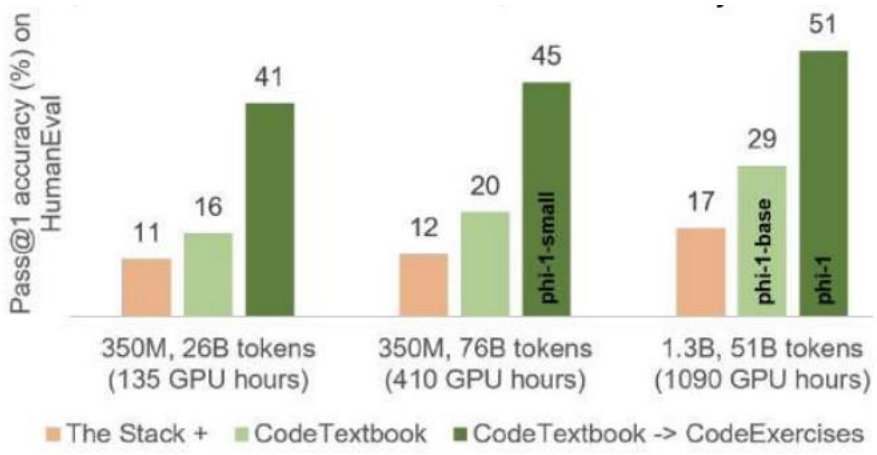

Рис. 1. Показатели эффективности небольших ИИ-моделей

- Предполагая, что текущие темпы потребления данных сохранятся, исследование Epoch AI предсказывает, что **человечество может исчерпать запас низкокачественных данных к 2030-2050 годам, высококачественных данных – к 2026 году**. Разработками, которые могут бросить вызов этой тенденции, являются **системы распознавания речи**, например, Whisper от OpenAI, которые могли бы дополнительно **сделать аудиоданные доступными для LLM**.

- **Создание новых белков** таким образом, чтобы они обладали желаемыми функциями или структурными свойствами, представляет очень **большой интерес для научных исследований и промышленности.** Ввиду успеха **диффузионных моделей** в генеративном моделировании образов и речи, ученые все чаще начинают применять их для разработки таких белков. RFdiffusion может генерировать белковые каркасы с желаемыми характеристиками, и ProteinMPNN затем используется для создания усовершенствованных последовательностей молекул.

- **В топ-20 научных областей, применяющих искусственный интеллект** для ускорения прогресса, входят **физические, социальные науки, науки о жизни и здоровье**. Из всех публикаций **наибольший прирост числа приходится на медицину**. В обозримом будущем в результате использования ИИ в естественных науках могут произойти наиболее значительные научные прорывы.

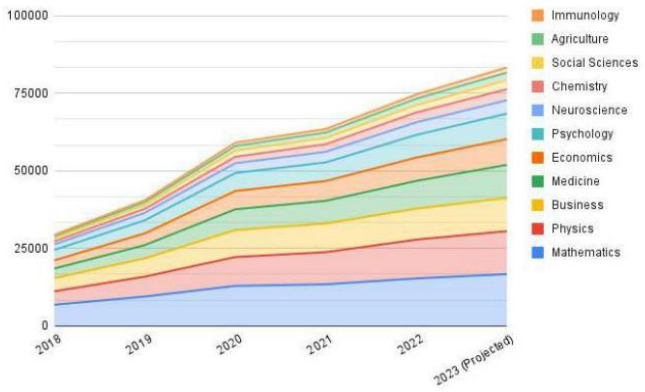

Рис. 2. Число научных публикаций о применении ИИ, по сферам, 2018-2022 гг.

- **>70% наиболее цитируемых статей по ИИ за последние 3 года написаны авторами из учреждений и организаций, базирующихся в США. За ними следует Китай и Великобритания**, однако разрыв с лидером является слишком значительным, чтобы расстановка сил изменилась в ближайшее время.

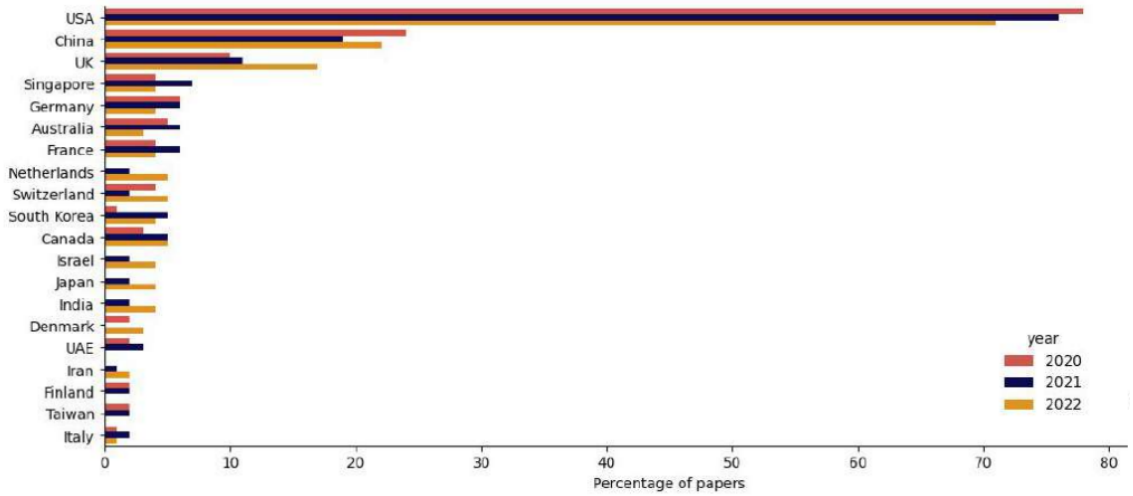

Рис. 3. Число научных публикаций по ИИ, по странам, 2020-2022 гг.

2. Промышленность

- Возросший **спрос на графические процессоры обеспечил NVIDIA рыночной капитализацией в 1 трлн долл. США, а их чипы используются в исследованиях в сфере ИИ в 19 раз чаще, чем все альтернативы вместе взятые.** Самым популярным из них стал **V100**, выпущенный в 2017 году, что показывает значительный срок службы разработок компании. «Войны чипов» также вынуждают отрасль адаптироваться к новым условиям: **NVIDIA, Intel и AMD стремятся создать специальные, совместимые с санкциями чипы для своих китайских клиентов**.

Рис. 4. Частота использования чипов различных компаний в исследовательских работах

- По сравнению с наиболее популярными действующими социальными сетями и мессенджерами, (YouTube, TikTok или WhatsApp), **приложения генеративного ИИ**, такие как ChatGPT, Runway или Character.ai страдают от **низкой медианы удержания и ежедневной активности пользователей**.

- **Окружной суд США подтвердил принцип, согласно которому только созданное человеком произведение может быть защищено авторскими правами**. В новом законе США говорится, что для любого художественного произведения необходим автор-человек и что компании должны указывать, где в их продуктах использовался искусственный интеллект. В связи с этим, **организации, предлагающие услуги по маркировке ИИ-контента**, такие как Scale AI и Surge HQ, **отмечают исключительный рост прибыли благодаря растущей популярности LLM**.

- Финансирование стартапов, использующих искусственный интеллект, в первом полугодии 2023 года было почти на уровне первого полугодия 2022 года, однако **без вливания капитала в генеративный ИИ общие инвестиции в эту технологию упали бы на 40% по сравнению с прошлым годом**. При этом инвестиционная активность частных компаний США и Великобритании стабильна, в то время как **капитал европейских ИИ-компаний упал более чем на 70%. США также продолжают лидировать по количеству ИИ-единорогов (315), за ними следуют Китай (70) и Великобритания (27).**

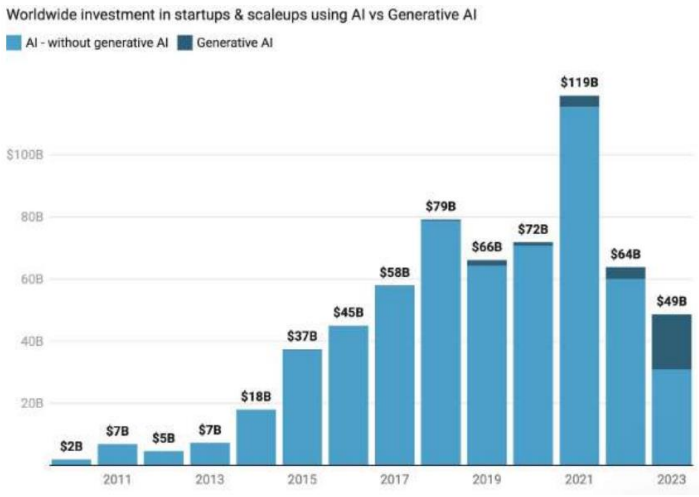

Рис. 5. Динамика инвестиций в стартапы, связанных с ИИ и генеративным ИИ, 2011-2023 гг.

- **На компании, работающие с генеративным ИИ, пришлось на 33% больше начального финансирования и на 130% инвестиций в раунде А, чем на все стартапы в 2023 году**. Общий объем капитала, вложенного в генеративные технологии, составил **18 млрд долл. США**, увеличившись почти в 5 раз с прошлого года (3,9 млрд долл. США). ChatGPT от OpenAI стал самым быстрорастущим интернет-продуктом в истории, а сама компания обеспечила себе прибыль в 1 млрд долл. США.

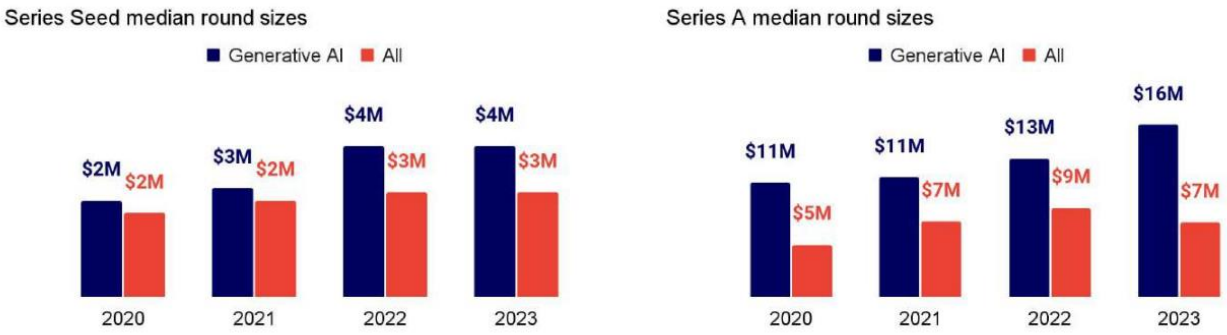

Рис. 6. Динамика инвестиционной активности в ИИ и генеративный ИИ, 2020-2023 гг.

3. Политика

- В 2023 году стало понятно, что **подходы к регулированию ИИ** стабилизируются и сводятся к нескольким траекториям:

1. **Подход незначительных изменений**. Представленный Великобританией и Индией, этот подход основан на том, что **искусственный интеллект в настоящее время не требует какого-либо дополнительного законодательства**.

2. **Инновационный подход**. Внедрение законодательных рамок, специфичных для ИИ, является основным подходом ЕС и Китая. К ним относятся оценка рисков, раскрытие информации о том,

когда контент генерируется искусственным интеллектом, предотвращение создания моделью незаконного контента и публикация информации о защищенных авторским правом данных, используемых для обучения.

3. **Стратегия запрета**. Обосновывая это соображениями безопасности, правительства ряда стран (Куба, Иран, Россия) **блокируют доступ к ИИ-инструментам**, в том числе к ChatGPT.

4. **Смешанный подход**. На других рынках мы наблюдаем либо **ослабление национального регулирования, либо принятие местных законов или урезанных правил для ИИ.** Канада пытается принять сокращенную версию Закона ЕС об искусственном интеллекте, запрещая некоторые приложения и регулируя другие. Вместо шкалы обязательств в стиле ЕС канадский закон об искусственном интеллекте и данных регулирует только приложения с высоким риском.

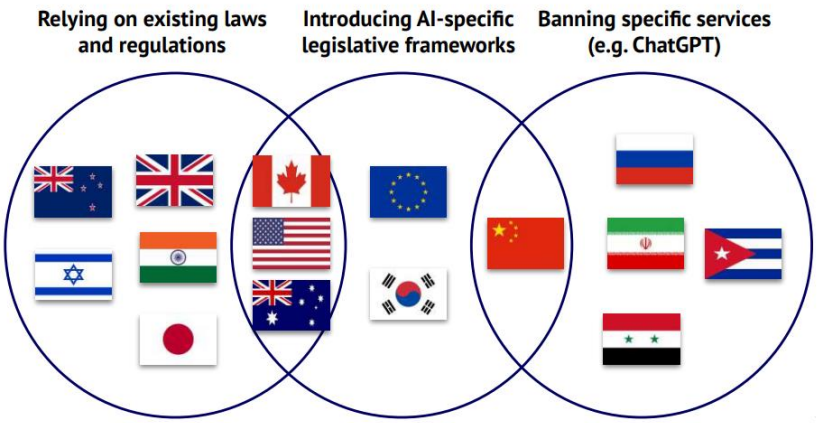

Рис. 7. Три подхода к регулированию ИИ, по странам

- В конце прошлого года **США ввели самый жесткий режим экспортного контроля в отношении Китая за последние десятилетия, запретив продажу продвинутых чипов** или инструментов, изготовленных для их использования, китайским фирмам. Это обозначило отказ от прежней политики, направленной на замедление китайского технологического процесса, в пользу активных попыток снизить китайский потенциал. При этом **критики задаются вопросом, насколько реалистично стремление США и ЕС к самообеспечению, учитывая отсутствие у них полного контроля над рынком полупроводников.**

- Такое обострившееся соперничество сопряжено с **рисками для обеих сторон**. Ограничивая доступ в Китай, **США рискуют ослабить рынок для своих собственных производителей** (что подрывает цель Закона о чипах), в то время как **Китай может продвинуть ограничения на металл только до тех пор, пока это не нанесет ущерб своим экспортерам**.

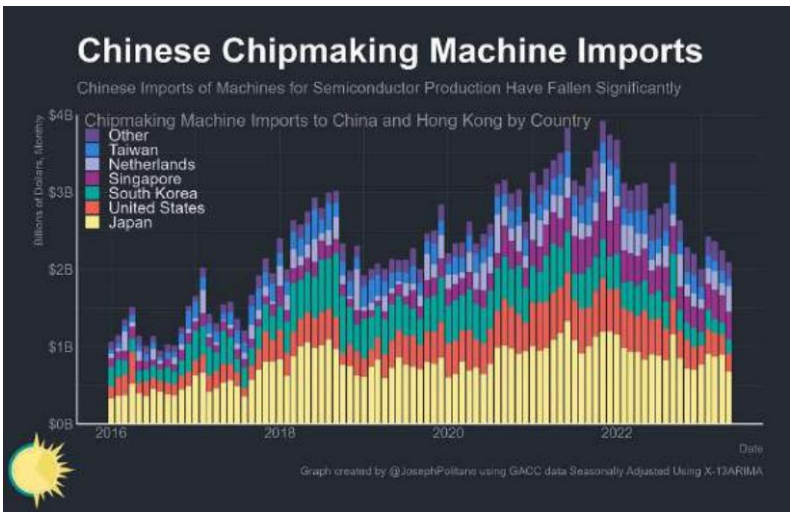

Рис. 8. Объем импорта чипов в Китай

4. Безопасность

- **2023 год стал годом дебатов о рисках ИИ, так как предполагаемые сроки создания сильного ИИ стали короче, чем прогнозировалось ранее.** В недавних выступлениях эксперты сосредоточились на опасениях, что автономные системы могут начать разрабатывать свои собственные системы манипуляции людьми и получения большего контроля, причиняя риск данным и жизни человека.

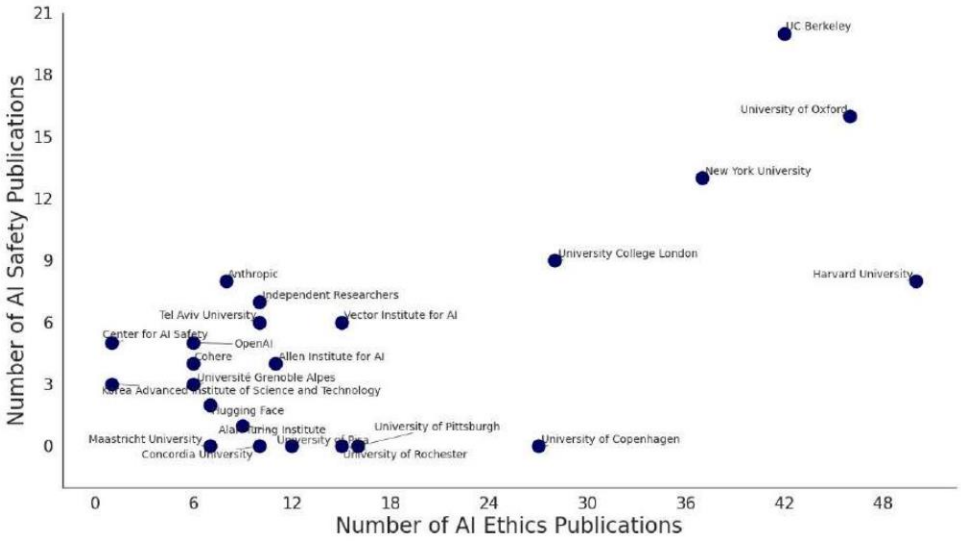

Рис. 9. Число публикаций на тему этики и безопасности ИИ

- До сих пор **остается легким реализация джейлбрейка** моделей ИИ даже за пределами API. Атаки, основанные на двух наиболее популярных принципах джейлбрейка, были успешны более чем в 96% проанализированных случаев.

- Чтобы смягчить описанные выше риски, **исследователи предлагают осуществлять предварительную тренировку ИИ с учетом новых запросов человека и «вознаграждая» ИИ-модель за правильное «поведение» отметкой «хорошо».** Они сообщают, что использование подобной техники может значительно уменьшить объем генерации нежелательного контента. Впоследствии, по мере того как модели становятся более функциональными и объемы выведенных данных значительно растут, одним из путей развития является **использование искусственного интеллекта для содействия обучению новых ИИ-моделей.**

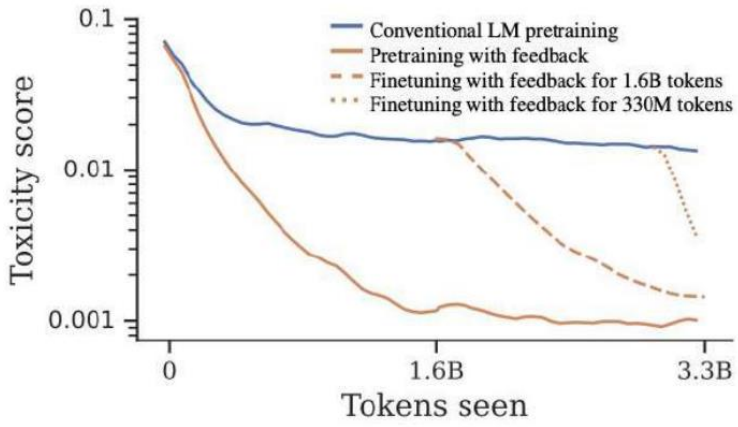

Рис. 10. Сравнение эффективности разных видов обучения ИИ-моделей

Методология отчета:

Ежегодное исследование, посвященное оценке текущего состояния ИИ в мире. В отчете авторы обращались к целому ряду партнеров, работающих в сфере ИИ. Выводы отчета строились на сравнениях с прошлыми годами, данных инсайдеров, а также на основании выходящих научных работ.

Применимость отчета в Российской Федерации:

Рекомендации из отчета применимы для обеспечения непрерывного развития сферы искусственного интеллекта в России:

- В контексте мировых исследовательских ИИ-трендов важно **продолжать создание новых больших языковых моделей**, не только увеличивая размеры их наборов данных, но и рассматривая **применение узконаправленных небольших моделей в различных отраслях**.

- В связи с **возможным дефицитом данных** в ближайшем будущем, **российским разработчикам важно создавать новые наборы**, извлекая их из фотографий и видео, а также **развивать сферу синтетических данных.**

- В связи с продолжающимся **противостоянием США и Китая** в сфере ИИ, **России важно правильно позиционировать себя в рамках этого соперничества**. Более того, в рамках развивающегося регулирования искусственного интеллекта, необходимо **проводить сравнения между стратегиями развития сферы ИИ в разных странах, адаптируя лучшие практики под российский контекст**.

- Проблема «джейлбрейка» больших языковых моделей, а также иных негативных аспектов ИИ (дипфейки, кража личных данных) все еще является актуальной, поэтому **государственным органам и компаниям-разработчикам важно сотрудничать, создавая новые способы нивелирования данных рисков**.