

МАРТ`23

01

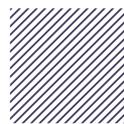
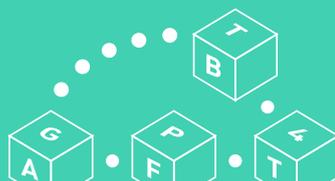


ФИНТЕХ -РАДАР

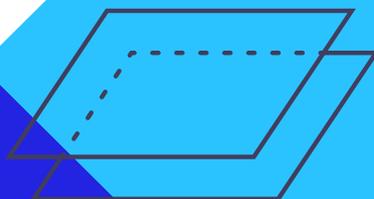
АФТ-инсайты
Техно-новости

ТЕХНОЛОГИЯ МЕСЯЦА:

GPT-4



 **ФИНТЕХ**
АССОЦИАЦИЯ



ЦИФРОВАЯ ЗРЕЛОСТЬ ФИНАНСОВОГО РЫНКА



Результаты исследования компаний финансового сектора в отношении ключевых критериев, которыми они руководствуются при оценке реализуемых программ цифровой трансформации.

ПО КАКИМ КРИТЕРИЯМ ВАША КОМПАНИЯ ОЦЕНИВАЕТ ЦИФРОВУЮ ЗРЕЛОСТЬ?



50% По операционной эффективности и наличию безбумажных процессов в бэк-офисе



28% По уровню использования данных в компании и их монетизации



42% По наличию новых цифровых и экосистемных продуктов



24% По зрелости цифровых каналов



41% По наличию платформенных решений и внедренных технологий



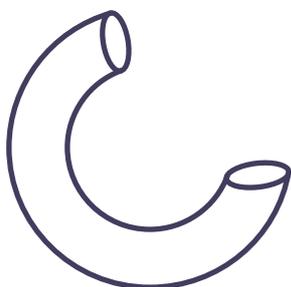
16% По уровню использования моделей искусственного интеллекта (ИИ)



30% По уровню защиты от киберугроз

Команда управления исследований и аналитики Ассоциации ФинТех (АФТ) провела опрос, который показал, что компании финансового сектора оценивают свою цифровую зрелость прежде всего по снижению издержек и повышению эффективности процессов. Драйвером цифровизации выступает наличие новых цифровых сервисов и успешный переход на безбумажный документооборот.

Несмотря на то, что все участники рынка тестируют или внедряют инструменты искусственного интеллекта, только 16% участников опроса оценивают зрелость своей компании по уровню использования моделей ИИ. Это, вероятно, связано с небольшой долей проектов с ИИ в промышленной эксплуатации и ограниченным применением решений на практике.



Одна из ключевых целей АФТ – предоставление независимой экспертизы за счет проведения аналитических исследований в области финансово-технологических трендов и бенчмаркинга цифровой зрелости участников рынка.

Для реализации этого Ассоциация разработала собственную методику определения уровня цифровой зрелости по 6 основным направлениям, а именно:

1. цифровые продукты и новые бизнес-модели,
2. каналы и клиентский опыт,
3. операционная эффективность,
4. данные как актив,
5. безопасность и риски,
6. информационные технологии.

В рамках каждого из них учитываются ключевые показатели эффективности. Результатом исследования становится радар цифровой зрелости компании, отражающий ключевые показатели эффективности по каждому из выделенных направлений, которые позволяют оценить уровень текущей цифровой зрелости и перспективные направления дальнейшего развития.





10 ТРЕНДОВ, КОТОРЫЕ ИЗМЕНЯТ ФИНАНСОВЫЙ РЫНОК

Команда исследователей АФТ провела опрос по наиболее значимым трендам развития финансовых технологий.

КАКИЕ ФИНАНСОВО-ТЕХНОЛОГИЧЕСКИЕ ТРЕНДЫ ВЫ СЧИТАЕТЕ КЛЮЧЕВЫМИ В 2023 ГОДУ?



37%

Открытые API



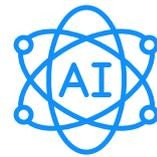
33%

Трансграничные платежи



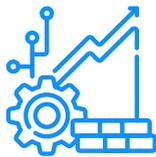
29%

Цифровые валюты центральных банков



26%

Вездесущий AI



25%

Цифровые финансовые активы и DeFi



25%

Развитие экосистем платежных сервисов



19%

Цифровое регулирование



15%

Устойчивые финансы



14%

Деглобализация

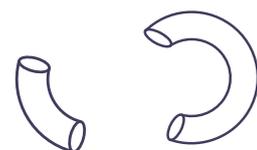


12%

Экосистемы идентификации

37% участников опроса выделили среди наиболее значимых в 2023 году трендов внедрение среды **открытых API**. Еще для 33% участников ключевым трендом является развитие **трансграничных платежей** и **платежных сервисов** для использования за границей.

С учетом общемирового тренда на выпуск **цифровых валют центральных банков**, для России важным трендом является тестирование и переход к использованию цифрового рубля. Данный тренд является общемировым, в том числе в рамках общей деглобализации. Например, Китай впервые начал тестирование цифровой валюты еще в 2020 году. Сегодня цифровой юань интегрирован во многие платежные сервисы, включая экосистемы WeChat и Tencent. Россия находится в глобальном тренде, запуская проект **цифрового рубля**, что позволит повысить уровень финансовой доступности и даст возможность создавать новые цифровые продукты и сервисы.



В 2023 году российский финансовый рынок по-прежнему будет фокусироваться на **развитии цифровых финансовых активов**, поиске практических бизнес-кейсов по токенизации финансовых инструментов в форме цифровых активов.

Использование технологий на основе решений искусственного интеллекта, **«вездесущий ИИ»** – еще один тренд, который выделили 26% участников. Участники российского финансового рынка продолжают развивать новые инструменты аналитики данных на основе технологий ИИ: адаптивные, самообучающиеся и способные самостоятельно создавать новый контент. Большое внимание уделяется инструментам, позволяющим снизить стоимость применения ИИ, которые создают условия для более глубокого проникновения ИИ в операционные процессы компаний и разработку новых продуктов.

КЛЮЧЕВЫЕ ТЕХНОЛОГИИ ЦИФРОВИЗАЦИИ В 2023 ГОДУ



Технология ИИ красной нитью проходит через все тренды, рассмотренные в исследовании [АФТ «3x10 трендов 2023 года»](#).

В результате исследования АФТ о цифровой зрелости в ипотеке, были выделены ключевые перспективные решения на основе искусственного интеллекта, на которые смотрят участники финансового рынка в 2023 году.

КЛЮЧЕВЫЕ ТЕХНОЛОГИИ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В 2023

■ Внедрено



83% Роботизация процессов (RPA)



72% Оптическое распознавание документов (OCR)



69% Текстовые чат-боты



62% Предиктивная аналитика и машинное обучение



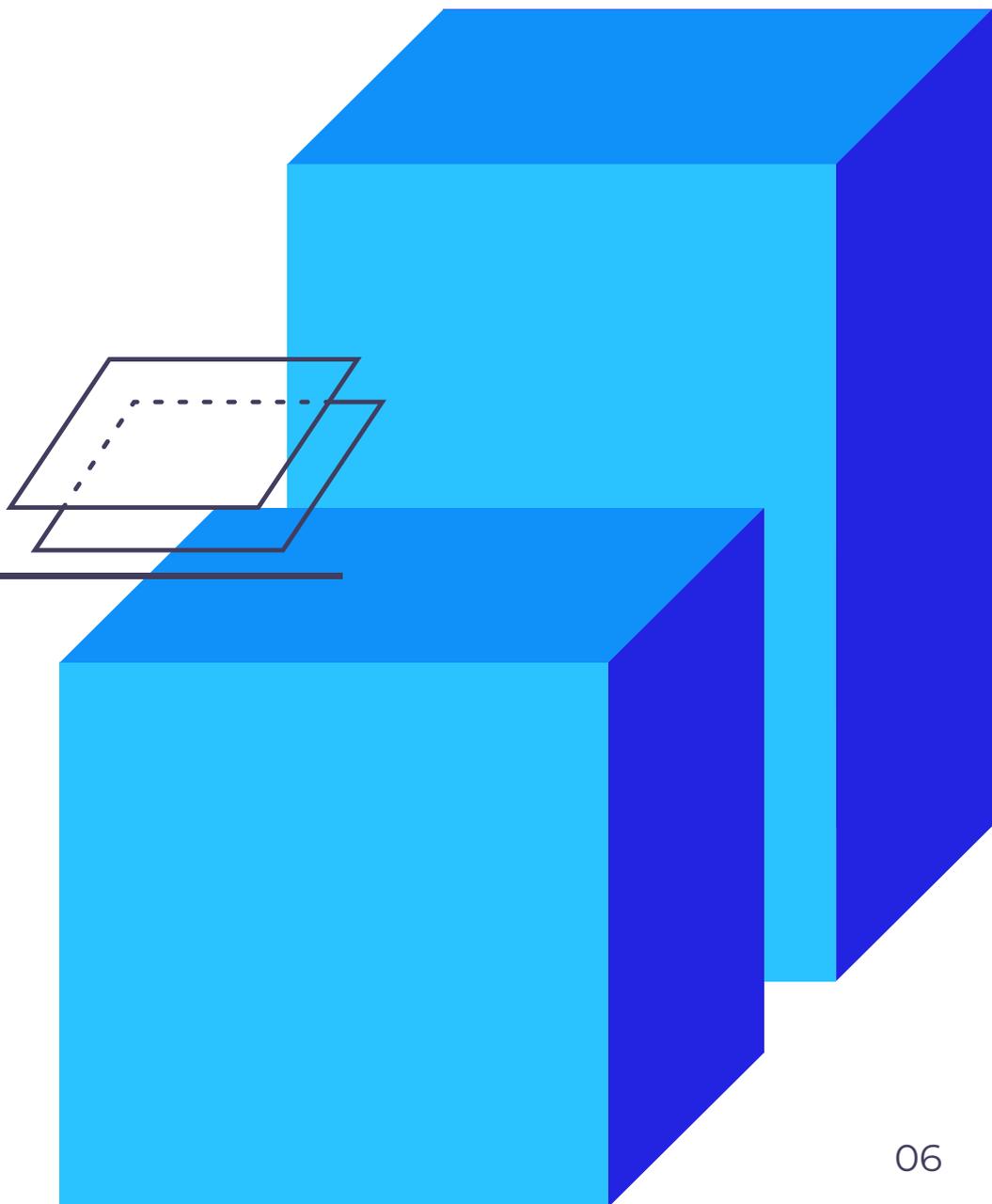
58% Голосовые ассистенты и речевые технологии

В 2023 году российские банки продолжают внедрять технологии **интеллектуальной роботизации процессов (IPA)**. 83% участников финансового рынка уже внедрили упомянутые технологий в свои процессы.

Помимо интеллектуальной роботизации процессов участники рынка активно внедряют технологии распознавания документов для перехода на безбумажный документооборот в рамках программ по цифровизации процессов. Также, согласно исследованию, участники активно используют **текстовых чат-ботов (69%), предиктивную аналитику и машинное обучение (62%).**

Дополнительно особый интерес сейчас вызывают модели **генеративного искусственного интеллекта**. Они отличаются от других моделей тем, что не только делают прогнозы или совершают действия на основе имеющихся данных, но и самостоятельно изучают их базовое распределение и используют эту информацию для создания новых синтетических данных.

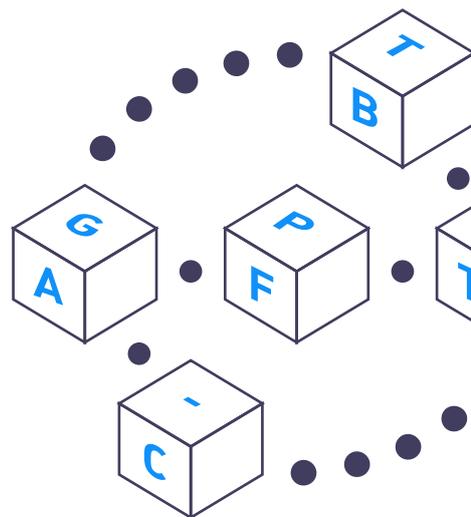
Ключевое различие между генеративными моделями и другими моделями ИИ заключается в том, что первые часто можно использовать для создания бесконечного числа новых выборок, тогда как другие модели могут делать прогнозы только для отдельных случаев. Хорошим примером генеративного ИИ служит недавно запущенная языковая модель **GPT-4**.



GPT-4



GPT-4 — (Generative Pre-trained Transformer 4) — это большая мультимодальная языковая модель компании OpenAI, которая генерирует текст на основе текстового фрагмента, введенного пользователем, или предложенного графического объекта. GPT-4 может решать более сложные задачи, чем предыдущие модели GPT. Она демонстрирует результаты на уровне человека во многих профессиональных и академических местах, включая международные олимпиады по биологии.



ПОЯВЛЕНИЕ И ЭВОЛЮЦИЯ ТЕХНОЛОГИИ

GPT-4 – технология, которую ученые приписывают к [«языковым моделям»](#) (Language Models). На самом деле, упрощенно, их основной функционал – угадывать, какое следующее слово должно идти за уже имеющимся текстом. Современные модели гораздо больше предшественников. Под большими моделями ([Large Language Models, LLM](#)) понимаются модели, имеющие огромное количество параметров, на сегодняшний день речь идет о сотнях миллиардов.

Как ранее было сказано, GPT расшифровывается как **Generative Pre-trained Transformer**, или «генеративный предобученный трансформер». [Трансформер](#) – это разработанный Google в 2017 году тип нейросетей¹, направленный на создание последовательностей с эффективной обработкой взаимосвязей между отдельными элементами. Например, словами в тексте. На сегодня это самая продвинутая техника в области обработки естественного языка (NLP).

Прорыв состоял в том, что был создан алгоритм, в котором не использовались привычные сверточные (CNN) и рекуррентные (RNN) нейронные сети, справлявшиеся с задачами лучше, чем иные решения на тот момент. Это достигается радикальным использованием [«механизма внимания»](#), который определяет, какая часть данных важнее другой, в зависимости от контекста. Взаимосвязи находятся не только среди введенных данных, но и тех, которые создает трансформер. Профессиональное сообщество оценило то, насколько хорошо передается информация между многочисленными этапами работы и функциональными блоками в архитектуре модели, насколько легко ее оптимизировать под конкретные задачи и насколько она эффективна в сравнении с аналогами.

Трансформеры до того легко масштабируются и универсальны, что во всех областях искусственного интеллекта начали активно их адаптировать и применять – от обработки текстов и изображений до звуков и видео, а также понимания устройства мира. Модель **GPT-1**, основанная на архитектуре трансформера, была выпущена в 2018 году и доказала,

¹Искусственная нейронная сеть — математическая модель, а также ее программные или аппаратные реализации, построенная в некотором смысле по образу сетей нервных клеток живого организма.



что нейросеть может генерировать тексты. У GPT-1 было 117 млн параметров¹. В 2019 году была выпущена модель **GPT-2**, которая превосходила предшественника по объему тренировочных данных и была в объеме больше в 10 раз (1,5 млрд параметров). К тому же она обучалась качественно новым навыкам, таким как написание связных эссе и решение сложных задач, требующих понимания устройства мира.

GPT-3 образца 2020 года была в 100 раз больше GPT-2 по количеству параметров (175 млрд) и в 10 раз – по объему тренировочных данных. Рост снова привел к значительному скачку в качестве: модель стала еще более универсальной, научилась переводу с других языков, арифметике, базовому программированию, пошаговым рассуждениям и прочим навыкам.

В ноябре 2022 года OpenAI сделало еще один интересный шаг: запустила чат-бот, позволяющий общаться с ИИ на основе GPT-3 – **ChatGPT**. Это стало очередным прорывом в развитии и популяризации технологии. С технической точки зрения значительных нововведений не проводилось. Однако появились удобный интерфейс взаимодействия и открытый публичный доступ. 23 марта OpenAI [добавил](#) поддержку плагинов для ChatGPT — расширений, которые интегрируют его со сторонними сервисами и позволяют получать доступ к актуальной информации. К примеру, добавили расширение Wolfram, которое позволяет пользоваться вычислительным, математическим и аналитическим функционалом Wolfram|Alpha¹ в реальном времени.

14 марта 2023 года вышла модель **GPT-4**. На данный момент она уже подключена к ChatGPT. OpenAI не раскрывает количество параметров в этой итерации, но преимущественно говорит о конкретных качественных изменениях. Как утверждают разработчики, она может принимать на вход изображения и текст, работает на «уровне компетенций человека» в различных профессиональных и академических тестах. Так, на экзаменах Uniform Bar Exam, LSAT, SAT Math и SAT Evidence-Based Reading & Writing GPT-4 входит в топ-10% сдавших экзамены³.

OpenAI в анонсе технологии отмечает, что при обычном диалоге разница между GPT-4 и предыдущими моделями может быть едва уловимой. Разница проявляется, когда сложность задачи достигает достаточного порога — GPT-4 более надежен, креативен и способен обрабатывать гораздо более тонкие инструкции.

+

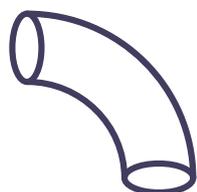
+

+

+

+

+



¹По количеству параметров принято оценивать размер модели LLM

²Wolfram Alpha - это вычислительная система знаний, которая может быть использована для ответа на широкий спектр вопросов и предоставления информации, основанной на данных.

³URL: <https://openai.com/research/gpt-4>.

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ

На данный момент всеобщего открытого доступа к GPT-4 нет, потому стоит переключиться на применение технологий, которые можно использовать уже сейчас, а именно ChatGPT.

ChatGPT способен выполнять следующий набор действий:

- генерировать и помогать улучшать разработку текстов и кода,
- резюмировать текст,
- классифицировать контент,
- отвечать на вопросы,
- переводить и преобразовывать язык (включая языки программирования).

Существует четыре основных способа применения технологии ChatGPT:

1. Как есть (AS-IS): ввод запросов и получение результатов через веб-интерфейс. На сегодняшний день это наиболее популярный подход.

2. Инжиниринг запросов без API¹: инжиниринг запросов – это использование сервиса, подобного ChatGPT, в сочетании с другими технологиями, как часть рабочего процесса. Этот рабочий процесс возможно создать вручную или с помощью технологий экранного скрейпинга и роботизированной автоматизации процессов (RPA).

3. Проектирование с использованием API: Используя API, можно напрямую общаться с технологией и более свободно использовать ее широкий функционал. К тому же только таким образом можно легально использовать ChatGPT в собственных приложениях. OpenAI дает возможность пользоваться API по тарифу \$0,002 за 1 тыс. токенов².

4. Пользовательская сборка: существует возможность под собственные нужды собрать новую генеративную модель, использующую GPT-3, GPT-2, их модификации или аналоги. Однако это будет уже не ChatGPT, поэтому будет теряться фильтрация запросов, взаимодействие с собеседником и другие полезные функции.



Ассоциация ФинТех недавно подготовила аналитическую справку по технологии GPT, и этот материал был создан с помощью самой технологии.

ИИ рассказал о себе сам.



¹API – Application Programming Interface, что значит программный интерфейс приложения. Интерфейс можно рассматривать как сервисный контракт между двумя приложениями. Этот контракт определяет, как они взаимодействуют друг с другом, используя запросы и ответы.

²Токены можно представить как фрагменты слов, прошедшие кодировку. Количество токенов, созданных для англоязычного текста, значительно меньше, чем для текстов на иных языках. Например, из «Ассоциация ФинТех» создастся 22 токена, а из «FinTech Association» - только 3. Таким образом, обработка названия организации на русском выйдет в 7 раз дороже, чем на английском.

КАКОВЫ ТЕКУЩИЕ ОГРАНИЧЕНИЯ?

- Нет гарантий конфиденциальности данных.
- Модель не может генерировать изображения, но в будущем технологию можно будет использовать в сочетании с визуальными генеративными моделями искусственного интеллекта.
- Полноценные протоколы безопасности в отношении ИИ еще не выработаны и не стандартизированы.

ВЛИЯНИЕ ТЕХНОЛОГИИ НА ОБЩЕСТВО

Согласно совместному исследованию OpenAI и экспертов Университета Пенсильвании ([GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models](#)), большие языковые модели, вроде GPT, могут затронуть до 80% рабочей силы в США, а 19% работников рискуют потерять от 50% рабочих задач. Из наиболее уязвимых профессий выделяются следующие: писатели, веб-дизайнеры, финансовые аналитики и ИТ-разработчики.

В отличие от предыдущих волн автоматизации, больше всего могут пострадать профессии с высоким уровнем дохода. Если применение технологий LLM и, в частности, GPT-4 будет направлено только на автоматизацию труда и замещение людей, это будет приводить к еще более сильной концентрации власти, богатства и ресурсов. У людей, которые на данный момент ими не обладают, не будет адекватной возможности улучшать свои результаты. Такой феномен Эрик Бринолфссон в статье [«The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence»](#) называет **«ловушкой Тьюринга»**.

Есть и более благоприятный сценарий развития событий: ИИ чат-боты станут мощными инструментами для работников, **дополняющими или «аугментирующими»** человеческую деятельность. С их помощью сотрудники компаний приобретут новые навыки и повысят собственную квалификацию. Этот феномен даст толчок развитию экономики в целом. Но данный сценарий предполагает целенаправленное усилие общества и государства на смену фокуса с автоматизации на аугментацию.

Опасения вокруг рынка труда вкупе с непроработанными протоколами безопасности в отношении ИИ вылились в петицию под названием [«Приостановить гигантские эксперименты с искусственным интеллектом: открытое письмо»](#), в которой содержится призыв приостановить как минимум на полгода обучение различных передовых систем ИИ. Обращение подписали Илон Маск, Стив Возняк и ещё около 3000 экспертов в области ИИ.



ИНЫЕ ЯЗЫКОВЫЕ МОДЕЛИ

Все модели GPT были созданы компанией OpenAI. Существуют и другие большие языковые модели, которые активно развивают мировые технологические лидеры. Так, помимо анонса GPT-4 14 марта, Google открыла сторонним разработчикам доступ к API нейросети PaLM.

В конце февраля 2023 компания Meta (организация запрещена на территории РФ) представила модель LLaMA. Как пишет [The Verge](#), LLaMA представляет собой не единую систему, а «квартет моделей» разного размера, которые будут доступны по некоммерческой лицензии, «ориентированной на исследовательские варианты использования». Аудитория пользователей — это университеты, НПО и отраслевые лаборатории.

Наиболее значимая для индустрии российская большая языковая модель на данный момент — **YaLM** от «Яндекса» на **100 млрд параметров**. Сейчас разрабатывается **YaLM 2.0** — новая версия генеративной сети. Предполагается, что она станет частью «Поиска», «Алисы», «Почты» и других сервисов компании. «Яндекс» уже частично использует генеративные нейронные сети при формировании поисковой выдачи, но сейчас у них скорее вспомогательная роль — они помогают сориентироваться, но не генерируют связные ответы.

Также Sber Cloud разработал собственную большую языковую модель **ruGPT-3**. На данный момент самая крупная ее версия ruGPT-3 XL имеет 1.3 млрд параметров.

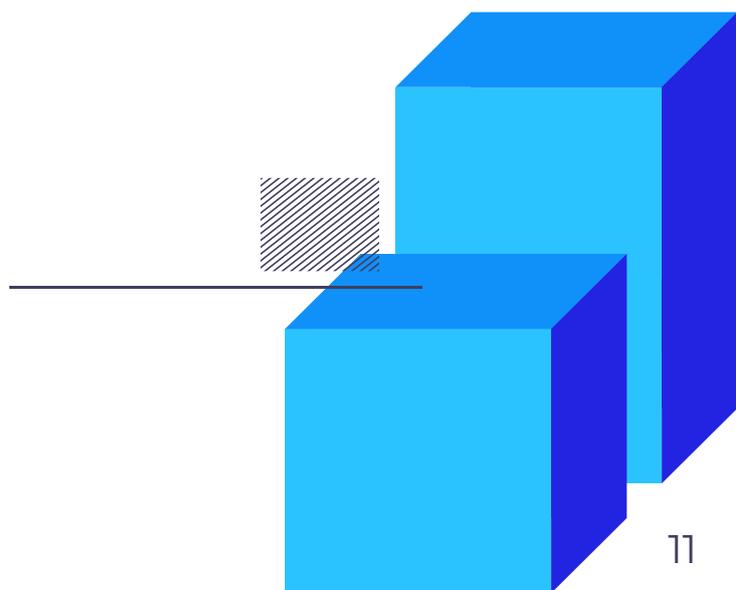
ЧТО ДЕЛАТЬ ДАЛЬШЕ?



Действовать разумно. Проект находится на очень ранней стадии и многое из того, что появляется в публичном пространстве, может оказаться субъективным мнением отдельного человека. Тем не менее, потенциал технологии значителен.

Изучить другие виды генеративного ИИ. Не стоит ограничиваться только моделями, генерирующими текст.

Экспериментировать. Необходимо изучить существующие угрозы и возможности этой технологии, составить дорожную карту для их выявления для конкретной компании, определить необходимые навыки и требуемые инвестиции.



ЗАКЛЮЧЕНИЕ

Искусственный интеллект уже заведомо включили в список потенциальных технологий общего назначения ([general purpose technology](#)), способных стать причиной технологической революции, радикально изменить общество и технологические уклады благодаря своему воздействию на существующие экономические и социальные структуры. К таким, например, относят паровой двигатель, электричество и компьютер.

С появлением GPT-4 и других современных LLM общественности стало понятно, что искусственный интеллект в каждом доме и устройстве – не только плод воображения писателей-фантастов, но довольно реалистичный сценарий развития человечества. Следить за этим со стороны – идея не лучшая, в адаптации к потенциальным радикальным изменениям необходимо принимать активное участие.

ПОЧИТАТЬ ДОПОЛНИТЕЛЬНО ПО ТЕМЕ:



BCG:

[Руководство CEO относительно революции в области генеративного ИИ](#)



IOT ANALYTICS:

[Отчет о тенденциях в области генеративного ИИ: общие сведения, тенденции, кейсы и возможные сценарии использования](#)



GARTNER:

[Помимо ChatGPT: Будущее генеративного ИИ для бизнеса. Кейсы, риски, иные технологии.](#)



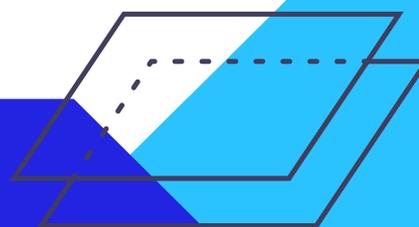
MCKINSEY:

[Генеративный искусственный интеллект уже здесь: как такие инструменты, как ChatGPT, могут изменить ваш бизнес](#)



ЭРИК БРИНОЛФССОН

[Ловушка Тьюринга: Перспективы и угрозы развития человекоподобного ИИ](#)





Техно-новости



GOOGLE ОТКРЫЛ ДОСТУП К API И ОБЛАКУ ДЛЯ СВОИХ МОДЕЛЕЙ PALM

PaLM API — это простая точка входа в большие языковые модели Google, которые могут быть использованы для различных приложений. Он предоставит разработчикам доступ к моделям, оптимизированным под различные сценарии использования, такие как создание контента и чат, а также к моделям общего назначения, оптимизированным для таких сценариев использования, как обобщение, классификация и многое другое. Наряду с API выпущен MakerSuite — инструмент, позволяющий разработчикам быстро и легко создавать прототипы.

ЦЕНТР ИССЛЕДОВАНИЙ ФУНКЦИОНАЛЬНЫХ МОДЕЛЕЙ ПРИ СТЭНФОРДЕ ПРЕДСТАВИЛ МОДЕЛЬ ALPACA 7B

Alpaca 7B — это адаптированная под инструкции версия LLaMA с 7 млрд параметров и работающая с text-davinci-003. Alpaca считается прорывом, так как может работать на одном графическом процессоре с практической производительностью, удивительно мала и дешева в воспроизведении, т. к. на разработку ушло меньше 600 долларов.

MICROSOFT АНОНСИРУЕТ COPILOT: БУДУЩЕЕ ДОКУМЕНТОВ OFFICE НА БАЗЕ ИИ

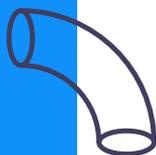
Компания Microsoft анонсирует Copilot с ИИ для своих приложений и сервисов Microsoft 365, призванный помочь людям в создании документов, электронных писем, презентаций и многого другого. Copilot, работающий на базе GPT-4 от OpenAI, будет работать в приложениях Microsoft 365 как помощник, появляясь на боковой панели в виде чат-бота, который позволит пользователям Office вызывать его для генерации текста в документах, создания презентаций PowerPoint на основе документов Word или даже для использования таких функций, как сводные таблицы (PivotTables) в Excel.

ГЕНЕРАТИВНАЯ НЕЙРОСЕТЬ MIDJOURNEY ДЛЯ СОЗДАНИЯ ИЗОБРАЖЕНИЙ ПО ТЕКСТОВОМУ ОПИСАНИЮ ОБНОВИЛАСЬ ДО ПЯТОЙ ВЕРСИИ

В пятой версии модель MidJourney научили создавать фотореалистичные изображения с большим количеством деталей. Некоторые пользователи уже начали делиться своими впечатлениями и изменениями, которые они заметили. Среди технических улучшений можно отметить более точное понимание запросов и увеличенную скорость создания изображений. Кроме этого, MidJourney открывает новые возможности для творчества. Дизайнеры утверждают, что нейросеть научилась «невероятно реалистично» передавать текстуру кожи и черты лица, создает более реалистичное и кинематографическое освещение, лучше передает отражения, блики и тени, выбирает более выразительные ракурсы и общий план.



КОМАНДА ИССЛЕДОВАНИЙ И АНАЛИТИКИ АФТ



МАРИАННА ДАНИЛИНА

Руководитель Управления исследований и аналитики

[E: m.danilina@fintechru.org](mailto:m.danilina@fintechru.org)



ДАРЬЯ ПЕТРОВА

Ведущий бизнес-аналитик по исследовательской деятельности

[E: d.petrova@fintechru.org](mailto:d.petrova@fintechru.org)



ГРИГОРИЙ КАРУНАС

Бизнес-аналитик по информационным сервисам

[E: g.karunas@fintechru.org](mailto:g.karunas@fintechru.org)



TELEGRAM КАНАЛ



WWW.FINTECHRU.ORG

