

# Аналитическая справка

по генеративным моделям искусственного интеллекта

Ассоциация ФинТех

Москва 2023



Генеративный Искусственный Интеллект (Generative AI) относится к области ИИ, направленной на создание новых и уникальных результатов, таких как изображения, текст, музыка и дизайн, с помощью алгоритмов машинного обучения. Генеративные модели ИИ обучаются на большом наборе примеров, которые могут поступать из различных источников. Процесс обучения включает использование комбинации алгоритмов машинного обучения и методов оптимизации для идентификации и изучения шаблонов и полезных структур данных. После того, как модель обучена, ее можно использовать для создания новых выходных данных путем выборки из изученного распределения вероятностей.

Генеративный ИИ – это быстро развивающаяся область, которая может изменить многое в нашей жизни, например процессы взаимодействия в бизнес-экосистеме.

Выпуск данной аналитической справки инициирован экспоненциальным прогрессом в области искусственного интеллекта и ростом интереса к вышеупомянутой технологии, что также нашло свое отражение в рамках исследования **«3x10 трендов на 2023 год»**, подготовленного Ассоциацией ФинТех, где детально рассматриваются ключевые тренды в парадигме **бизнес, технологических и финтех-трендов**, и при работе над которым стало очевидно, что прогресс в области искусственного интеллекта стал драйвером и катализатором развития других трендов.

***ВНИМАНИЕ:** Настоящая аналитическая справка подготовлена с применением модели генеративного искусственного интеллекта – ChatGPT.*



# Содержание

<b>1. История искусственного интеллекта</b>	<b>4</b>
- Ключевые вехи в эволюции моделей ИИ	4
- Характеристики моделей ИИ	5
<b>2. Генеративный искусственный интеллект</b>	<b>6</b>
- Прогресс в генеративных моделях	6
- Направления применения	7
- Владельцы инструментов	9
- Бизнес-модели для коммерциализации технологий	9
- Риски и вызовы	10
- Этические проблемы использования	10
- Применение в российском финансовом секторе	11



# 1. История ИИ

Историю искусственного интеллекта можно проследить до 1950-х годов, когда исследователи впервые начали изучать концепцию создания машин, способных выполнять задачи, для которых обычно требуется человеческий интеллект.

Область исследований ИИ была официально основана на конференции в Дартмутском колледже в 1956 году, где был придуман термин «искусственный интеллект». В последующие десятилетия искусственный интеллект добился значительного прогресса, включая разработку экспертных систем, способных выполнять задачи в специализированных областях, таких как медицина и инженерия, а также создание алгоритмов машинного обучения, позволяющих компьютерам учиться на основе данных.

Однако ИИ также столкнулся с неудачами, такими как событие под названием «зима ИИ» – период сокращения финансирования и интереса к этой области в 1970-х и 1980-х годах. В последние годы достижения в области машинного обучения, особенно глубокого обучения, привели к быстрому повышению производительности ИИ, что сделало его важной областью исследований и разработок.

## Ключевые вехи в эволюции моделей ИИ

Процесс эволюции технологий шел по пути увеличения сложности и возможностей, обусловленных достижениями в области компьютерного оборудования, алгоритмов и доступности данных:

- 1. Ранние модели ИИ** (1950–1960-е годы) – основывались на системах, основанных на правилах, и экспертных системах, которые использовали набор predetermined правил для принятия решений. Эти модели были ограничены в своих возможностях и могли справляться только с простыми задачами.
- 2. Статистические модели ИИ** (1960–1980-е годы) – основывались на статистических методах и могли решать более сложные задачи, такие как регрессия и классификация. Примером таких моделей служат линейная регрессия и деревья решений.
- 3. Нейронные сети** (1980–2010-е годы) – используют слои взаимосвязанных узлов для прогнозирования и способны решать сложные задачи, такие как распознавание изображений и речи.
- 4. Глубокое обучение** (с 2010-х по настоящее время) – тип нейронной сети с несколькими слоями, справляющаяся с широким спектром задач, включая распознавание изображений и речи, обработку естественного языка и геймификация, и достигшая «сверхчеловеческой» производительности во многих тестах.
- 5. Модели генеративного ИИ** (с 2010-х по настоящее время) – способны генерировать новые данные, которые напоминают данные, на которых они обучались.



## Характеристики моделей ИИ

Модели искусственного интеллекта имеют несколько важных характеристик, определяющих их производительность, возможности и ограничения:

**Точность** – характеристика означает, насколько хорошо модель ИИ может предсказать правильный результат для заданного ввода. Это мера производительности модели и часто выражается в процентах или в виде матрицы ошибок.

**Сложность** – относится к количеству параметров, нейронов или слоев в модели ИИ и определяет способность модели справляться со сложными задачами. Как правило, более сложные модели способны справляться с более сложными задачами, но они также более затратны в вычислительном отношении и склонны к переобучению.

**Обучающие данные** – это входные данные, используемые для обучения модели ИИ, и они являются критическим фактором производительности модели. Качество и количество обучающих данных, а также их разнообразие и представление играют роль в определении точности и надежности модели ИИ.

**Генерализация** – относится к способности модели ИИ хорошо работать с новыми, ранее неизвестными данными. Модель, которая хорошо обучена, имеет низкую дисперсию (low variance) и высокое смещение (high bias), в то время как модель, которая отличается переобучением, имеет высокую дисперсию (high variance) и низкое смещение (low bias).

**Объяснимость** – относится к способности понимать и интерпретировать решения и прогнозы модели ИИ. Некоторые модели ИИ, такие как деревья решений и линейная регрессия, более прозрачны и просты для интерпретации, в то время как другие, такие как глубокие нейронные сети, могут быть трудными для понимания и интерпретации.

**Смещение (предвзятость)** – относится к систематическим ошибкам в прогнозах модели ИИ и может быть результатом выбора смещенных обучающих данных или внутренних ограничений модели. Это может иметь серьезные последствия, особенно в таких областях, как здравоохранение, финансы и уголовное правосудие, поэтому важно выявлять и устранять предвзятость в моделях ИИ.



## 2. Генеративный ИИ

Особенный интерес в последние годы вызывают **модели генеративного искусственного интеллекта**. Они отличаются от других моделей, которые обучены делать прогнозы на основе входных данных или обучаются методом проб и ошибок, тем, что такие модели изучают базовое распределение данных и используют эту информацию для создания новых синтетических данных, а не делают прогнозы или предпринимают действия на основе имеющихся данных.

Еще одно 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 запущенный и уже нашумевший чат-бот **ChatGPT**. Технология представляет собой языковую модель искусственного интеллекта, разработанную компанией OpenAI. Он предназначен для генерации текста на основе предоставленного ему ввода. ChatGPT обучен работе с разнообразным интернет-текстом, включая разговорные данные из различных источников, и предназначен для ответа на широкий круг вопросов и предоставления актуальной информации. ChatGPT спроектирован так, чтобы быть гибким и адаптивным, что позволяет ему обрабатывать широкий спектр тем и стилей разговора. Его можно использовать для различных приложений, включая поддержку клиентов, чат-боты и языковой перевод.

### Прогресс в генеративных моделях ИИ

**GPT-3 (Generative Pretrained Transformer 3) и GPT-2 (Generative Pretrained Transformer 2)** – это языковые модели, разработанные американской компанией OpenAI. Модели схожи, но между ними есть некоторые ключевые отличия:

**Объем данных:** GPT-3 значительно больше, чем GPT-2. GPT-3 имеет более 175 миллиардов параметров, тогда как GPT-2 имеет только 1,5 миллиарда параметров. Преимущество в размере позволяет GPT-3 гораздо глубже понимать язык и генерировать более сложные выходные данные.

**Данные для обучения:** GPT-3 был обучен на гораздо большем наборе данных, чем GPT-2, включая информацию из самых разных источников, включая книги, статьи, веб-сайты и многое другое. Эти обучающие данные обеспечивают GPT-3 более полное понимание мира и возможность генерировать более точные и релевантные результаты.

**Возможности:** GPT-3 способен выполнять гораздо более широкий спектр задач, чем GPT-2, включая обработку естественного языка, машинный перевод, суммирование и даже кодирование. GPT-3 может генерировать очень связные с точки зрения грамматики предложения, что делает его более полезным для более широкого круга кейсов применения.

**Доступность:** GPT-3 не так широко распространен, как GPT-2, поскольку работает через частный API, для доступа к которому требуется регистрация и оплата. Напротив, GPT-2 более широко распространен и используется исследователями и разработчиками по всему миру для создания новых языковых моделей и приложений.



В целом GPT-3 представляет собой значительный шаг вперед по сравнению с GPT-2 с точки зрения объема обучающих данных, возможностей и производительности. Однако его ограниченная доступность и более высокая стоимость могут снизить его использование в ближайшем будущем.

## Направления применения генеративного ИИ

Генеративный ИИ может применяться в ряде областей, которых включают:

**Искусство:** генеративный ИИ можно использовать для создания новых произведений искусства, музыки, литературы и пр. В качестве примера можно привести основанный на глубоком обучении алгоритм генерации изображений **DRAW**, разработанный Google DeepMind. Сеть генерирует высококачественные изображения с мелкими деталями, такими как текст и текстуры, сохраняя при этом глобальную структуру, соответствующую входному изображению.

**Дизайн продукта:** инструменты генеративного дизайна на базе искусственного интеллекта могут помочь разработчикам быстро и эффективно создавать новые продукты, создавая несколько вариантов дизайна и предлагая лучший из них. Так, компания **Dell** использовала генеративный ИИ для разработки новых моделей своих ПК. Система искусственного интеллекта была обучена на данных предыдущих успешных продуктов и смогла создать новые конструкции, оптимизированные с точки зрения производительности, эффективности и стоимости.

**Здравоохранение:** генеративный ИИ можно использовать для создания новых методов лечения и разработки лекарств, анализа огромного количества данных пациентов для поиска корреляций и новых идей, а также для создания персонализированных планов лечения. Например, компания **Insilico Medicine** использовала генеративный ИИ для ускорения процесса разработки новых лекарств. Специалисты разработали платформу искусственного интеллекта, способную генерировать новые химические соединения, которые можно использовать в качестве лекарств, и успешно использовали эту технологию для разработки новых лекарств от таких заболеваний, как болезнь Паркинсона.

**Производство:** генеративный ИИ можно использовать для разработки и оптимизации сложных производственных процессов, например, используемых при производстве компонентов для аэрокосмической отрасли, автомобильных запчастей и т. д. Например, **Boeing** использовала генеративный ИИ для разработки деталей для своих самолетов, что привело к созданию более легких и эффективных компонентов. Система смогла создать конструкции, которые были на 40% легче исходных компонентов, что позволило снизить вес самолета и повысить эффективность использования топлива.

**Финансы:** генеративные системы на базе искусственного интеллекта можно использовать для создания финансовых прогнозов, торговых рекомендаций, новых инвестиционных стратегий. Например, испанский банк **BBVA** использовал генеративный ИИ для автоматизации обработки заявок на кредит, сократив время, необходимое для принятия решения о кредите, с нескольких дней до нескольких часов.

**Архитектура и строительство:** генеративный ИИ можно использовать для проектирования зданий и городов, оптимизируя такие факторы, как энергоэффективность, стоимость проектирования и устойчивость. Так, компания **Procore Technologies** использовала технологию для контроля качества для обнаружения дефектов и отклонений от проектных спецификаций в строительных проектах. ИИ-платформа была обучена на данных о предыдущих дефектах и смогла с высокой точностью выявлять новые случаи отклонений, повышая качество проектов.

**Обработка естественного языка:** генеративный ИИ можно использовать для создания новых текстов, резюме, переводов, ответов на вопросы. Технология применяется у бигтехов *Google, Microsoft, IBM* и пр.



В качестве дополнительного примера применения модели генеративного ИИ можно привести виртуального персонального помощника **Replika**. Этот проект, использующий ChatGPT, привлек большое внимание СМИ за свою способность оказывать поддержку в области психического здоровья. Replika предназначена для предоставления пользователям безопасного и непредвзятого пространства для обсуждения своих мыслей и чувств, а модель обучена реагировать эмпатично и предоставлять поддержку пользователям. Replika может помочь пользователям с широким спектром задач, включая поддержку психического здоровья, помощь пользователям в управлении их расписаниями и напоминаниями, а также предоставлять развлекательную активность и вести беседы.

Использование генеративного ИИ также набирает обороты в образовании. Например, в Уортонской школе при Университете Пенсильвании чат-бот GPT-3 смог **сдать выпускной экзамен** в рамках программы магистра делового администрирования (MBA). По словам профессора университета Кристиан Тервиша, бот получил на экзамене оценку «хорошо», «отлично справившись с основными вопросами управления операциями и анализа процессов, включая те, которые основаны на тематических исследованиях».

Также нашумевшим примером использования ChatGPT в образовании стало написание **студентом российского университета РГГУ** дипломной работы исключительно с использованием искусственного интеллекта. Молодой человек попросил нейронную сеть составить план дипломной работы, и на его основании написать текст документа. Таким образом, студент «написал» свой диплом за 23 часа. В рамках проверки работы на антиплагиат, оригинальность документа составила 82%.

В Голландии благодаря сотрудничеству ING Bank, J. Walter Thompson Amsterdam и Технического университета Делфта был создан арт-проект **The Next Rembrandt**, целью которого является создание нового портрета в стиле известного голландского художника Рембрандта Харменса ван Рейна с использованием передовых технологий, таких как 3D-печать и генеративный ИИ. Команда сначала отсканировала существующие картины Рембрандта, чтобы создать цифровой набор данных его уникального стиля живописи. Затем они использовали алгоритмы глубокого обучения для анализа этого набора данных и создания нового портрета в стиле Рембрандта.

В качестве дополнительного примера можно привести продукт **DeepFaceLab** – программный инструмент для создания дипфейков. Он использует искусственный интеллект и машинное обучение, чтобы заменить лицо одного человека в видео лицом другого человека, создавая дипфейк. DeepFaceLab использует глубокую нейронную сеть для анализа исходного видео и создания нового видео, в котором исходное лицо плавно заменяется целевым лицом. Инструмент используется как профессиональными цифровыми художниками, так и любителями для различных целей, таких как развлечения, исследования и образование.

Кроме того, компания OpenAI разработала нейросеть **DALL·E**, способную генерировать высококачественные изображения из текстовых описаний. Технология работает с использованием глубокой нейронной сети, обученной на большом наборе изображений и текстовых описаний. При наличии текстового описания изображения DALL·E создает изображение, соответствующее описанию. Система способна генерировать широкий спектр изображений, от фотореалистичных объектов до фантастических существ и абстрактных сцен.



Примером использования генеративного ИИ в музыкальной деятельности может служить музыкальная платформа, **Generative.fm**, использующая комбинацию машинного обучения и генеративных алгоритмов для создания музыки. Технология позволяет пользователям вводить определенные параметры, такие как жанр, настроение и темп, для создания музыки, которая соответствует их желаемым критериям. Платформа также предлагает широкий спектр возможностей настройки, таких как настройка сложности музыки и управление инструментами, используемыми в композиции.

Разработчик программного обеспечения на базе ИИ Latitude разработал текстовую приключенческую игру **AI Dungeon**, в которой искусственный интеллект создает активность в режиме реального времени. Используя AI Dungeon игроки могут вводить свой собственный выбор действий, и ИИ будет генерировать соответствующую историю на основе этих входных данных. ИИ использует глубокую нейронную сеть, обученную на большом наборе данных текстовых приключений, что позволяет ему генерировать широкий спектр различных историй и сценариев.

## Владельцы инструментов Генеративного ИИ

Право собственности на генеративные модели ИИ может зависеть от конкретных обстоятельств и контекста, в которых они создаются.

В некоторых случаях создатели модели ИИ могут владеть правами интеллектуальной собственности на модель, включая любые патенты, авторские права и коммерческие тайны. В других случаях модель ИИ может принадлежать организации, которая финансировала ее разработку, например корпорации или исследовательскому институту.

Если модель ИИ создана сотрудником компании, компания может владеть правами на модель в соответствии с доктриной «работы, сделанные по найму».

Когда дело доходит до моделей ИИ с открытым исходным кодом, владение моделью может быть более сложным, поскольку в ее разработку могли вносить свой вклад несколько участников. В этих случаях конкретные условия лицензии с открытым исходным кодом, в соответствии с которой выпущена модель, будут определять, кто владеет правами на модель и как ее можно использовать.

В целом, важно тщательно рассмотреть права собственности на модели ИИ и иметь четкие соглашения относительно их использования, особенно когда они разрабатываются для коммерческих целей.

## Бизнес-модели для коммерциализации технологий генеративного ИИ

Технологии генеративного искусственного интеллекта могут быть коммерциализованы с использованием различных бизнес-моделей, в том числе:

**Программное обеспечение как услуга (SaaS):** компания может предлагать доступ к своей технологии генеративного искусственного интеллекта в виде облачной услуги, при этом клиенты платят регулярную плату за доступ к технологии и ее возможностям.

**Лицензирование:** компания может лицензировать свою технологию генеративного искусственного интеллекта другим организациям, что позволяет им использовать эту технологию в своих собственных системах и процессах.



**Консалтинг:** компания может предлагать консультационные услуги, чтобы помочь организациям внедрить технологию генеративного ИИ в свою деятельность, предоставляя экспертные знания и поддержку по мере необходимости.

**Партнерство:** компания может сотрудничать с другими организациями для совместной разработки и коммерциализации генеративной технологии искусственного интеллекта, разделяя риски и выгоды от предприятия.

**Продукты:** Компания может разрабатывать и продавать генеративные продукты на основе ИИ, такие как творческие инструменты для художников, программное обеспечение для проектирования для инженеров или аналитические инструменты для медицинских работников.

Конкретная используемая бизнес-модель может зависеть от конкретных потребностей и целей компании, а также от рыночного спроса на ее продукты и услуги.

## Риски и вызовы при работе с моделями генеративного ИИ

Генеративные модели ИИ, как и любая другая технология ИИ, имеют как преимущества, так и риски, связанные с их использованием. Некоторые из рисков и проблем генеративного ИИ включают в себя:

**Предвзятость** – модели генеративного ИИ могут изучать и усиливать предубеждения, присутствующие в обучающих данных, что приводит к предвзятым результатам, которые могут укоренять стереотипы и усиливать дискриминацию.

**Неправильное использование** – использование ИИ для создания фейкового или вводящего в заблуждение контента, такого как фейки, дипфейки и синтетические медиа, которые занимаются дезинформацией общества.

**Риски конфиденциальности** – модели могут генерировать новые данные, которые напоминают реальных людей, или используются для связывания данных из разных источников для повторной идентификации людей.

**Слабый уровень регулирования** – в настоящее время существует всего несколько правил и стандартов, регулирующих разработку и развертывание генеративных моделей ИИ, и существует потребность в структуре, обеспечивающей ответственное использование этих моделей. Также распределение ответственности за действия и решения генеративных моделей ИИ пока находятся в «серой зоне».

**Технические ограничения** – генеративные модели ИИ все еще являются относительно новой технологией, и существуют технические ограничения и проблемы, связанные с их разработкой и развертыванием, такие как сложность обучения больших сложных моделей и риск сбоя системы, когда модель только генерирует ограниченное количество выходных данных.

## Этические проблемы применения моделей генеративного ИИ

Разработка и развертывание генеративных моделей ИИ поднимают важные этические вопросы, и крайне важно придерживаться этического подхода при работе с этими моделями. Среди ключевых принципов этического подхода к работе с генеративным ИИ можно выделить следующие:



1. Важно брать **ответственность** за влияние генеративных моделей ИИ и обеспечить их разработку и использование в соответствии с этическими и моральными ценностями.

2. Генеративные модели ИИ должны быть **прозрачными** в своей работе, и должна быть возможность понять, как модель генерирует свои результаты. Это помогает гарантировать, что модель несет ответственность за свои решения и устраняет любые предубеждения или непредвиденные последствия.

3. Генеративные модели ИИ следует разрабатывать и использовать таким образом, чтобы способствовать **справедливости** и предотвращать дискриминацию. Это включает в себя рассмотрение влияния модели на различные группы населения и принятие мер для смягчения любых погрешностей в обучающих данных.

4. Модели генеративного ИИ должны обеспечивать **конфиденциальность** и защищать **конфиденциальные данные**. Это включает в себя принятие мер для предотвращения неверной идентификации лиц и обеспечения того, чтобы сгенерированные данные не нарушали конфиденциальность.

5. Необходимо создание **структуры**, регулирующей разработку и развертывание генеративных моделей ИИ и обеспечивающая их **ответственное использование**. Это может включать разработку руководств и передовой практики, а также создание независимого надзорного органа для обеспечения соблюдения требований.

В целом, этический подход к работе с генеративным ИИ требует приверженности ответственному и этичному использованию, а также готовности участвовать в более широких социальных последствиях этих моделей искусственного интеллекта.

## Применение генеративного ИИ в российском финансовом секторе

Генеративный ИИ имеет ряд потенциальных применений в сфере финансовых услуг в России, в том числе:

**1. Управление инвестициями.** Генеративные модели ИИ можно использовать для создания новых инвестиционных стратегий и портфелей, а также для анализа финансовых данных и определения инвестиционных возможностей.

**2. Обнаружение мошенничества:** модели генеративного ИИ можно использовать для обнаружения и предотвращения финансового мошенничества, такого как отмывание денег и кража личных данных, путем анализа финансовых транзакций и выявления подозрительных моделей.

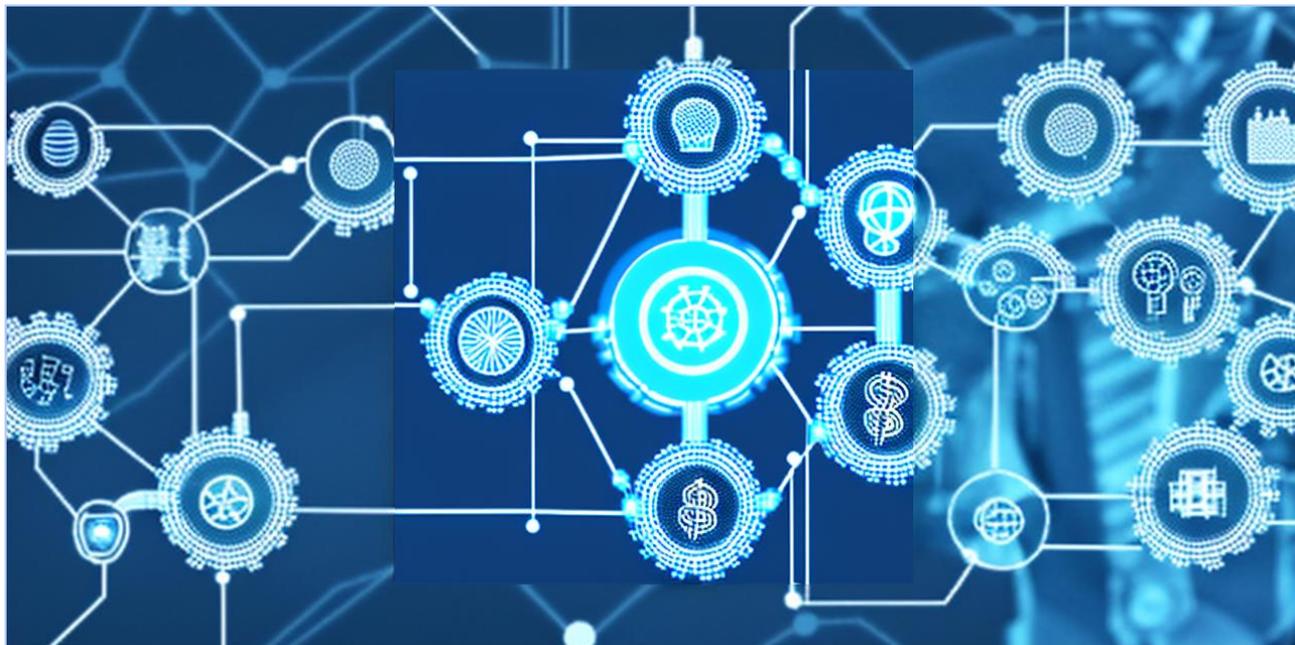
**3. Сегментация и персонализация клиентов.** Генеративные модели искусственного интеллекта можно использовать для создания пользовательских финансовых продуктов и услуг для определенных сегментов клиентов, что позволяет предоставлять более целенаправленные и персонализированные финансовые услуги.

**4. Анализ кредитного риска:** модели генеративного искусственного интеллекта можно использовать для анализа кредитного риска и принятия кредитных решений, таких как одобрение или отказ в кредите, путем анализа финансовых данных и кредитной истории.

**5. Соответствие регуляторным требованиям:** генеративные модели ИИ можно использовать для соблюдения финансовых правил, таких как правила борьбы с отмыванием денег (AML) и правила «знай своего клиента» (KYC), путем анализа финансовых данных и обнаружения подозрительных моделей.



Это всего лишь несколько примеров потенциального применения генеративного ИИ в сфере финансовых услуг. Поскольку технология продолжает развиваться и совершенствоваться, вполне вероятно, что в будущем будут разработаны новые и инновационные способы использования генеративного ИИ.



*Изображение создано моделью Stable Diffusion v1.5 по запросу «Generative AI for Finance industry»*

