ЛЕКЦИЯ 11
ТЕХНОЛОГИИ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Цель лекции

Сформировать системное представление о генеративном искусственном интеллекте, принципах работы основных классов генеративных моделей, технологиях генерации текста, изображений, программного кода, аудио и видео; рассмотреть области применения, способы оценки результатов, ограничения, риски и требования безопасного и ответственного использования генеративных систем.

Понятие генеративного искусственного интеллекта

Генеративный искусственный интеллект (Generative AI) — направление искусственного интеллекта, в котором модели создают новый контент на основе закономерностей, выявленных в обучающих данных. В отличие от классификатора, который выбирает один из известных классов, генеративная модель формирует новый объект: текст, изображение, программный код, аудиосигнал, видеопоследовательность или другое представление данных.

Генеративная система не является обычным хранилищем заранее подготовленных ответов. Результат формируется на основе параметров модели и текущего входного условия. Это условие может быть текстовым запросом, изображением, фрагментом аудио, кодом, структурированными данными или сочетанием нескольких типов информации.



Рисунок 1 – Основные типы контента, создаваемого генеративным ИИ

Генеративные и дискриминативные модели

Дискриминативные модели в первую очередь решают задачу различения объектов или прогнозирования целевого значения. Генеративные модели стремятся моделировать структуру данных так, чтобы создавать новые примеры, похожие по статистическим свойствам на обучающую выборку.

Характеристика	**Дискриминативная модель**	**Генеративная модель**
Основная цель	Определить класс или числовое значение	Создать новый объект или продолжение данных
Пример задачи	phishing / legitimate	Сформировать текст письма по условию
Тип результата	Метка, вероятность, числовой прогноз	Текст, изображение, аудио, видео, код
Что изучает	Границу или зависимость между входом и целью	Распределение и структуру данных
Пример	Логистическая регрессия, классификатор CNN	VAE, GAN, диффузионная или авторегрессионная модель

Что значит «генерировать» данные

Генерация не означает случайное создание любого объекта. Модель формирует результат в соответствии с закономерностями обучающих данных и заданными условиями. Поэтому качество генерации зависит от состава обучающего набора, архитектуры модели, процедуры обучения и способа управления генерацией.

Например, модель генерации изображений получает текстовое описание и создаёт визуальный результат, соответствующий этому описанию. Языковая модель получает последовательность токенов и формирует продолжение, которое статистически согласуется с контекстом.

Основные классы генеративных моделей

Класс модели	**Основная идея**	**Типичные данные**
Авторегрессионные модели	Генерируют следующий элемент на основе предыдущих	Текст, код, аудио, изображения
Автокодировщики и VAE	Изучают компактное скрытое представление и восстанавливают данные	Изображения, сигналы
GAN	Генератор и дискриминатор обучаются в состязательном процессе	Изображения и другие данные
Диффузионные модели	Обучаются восстанавливать данные из зашумленного состояния	Изображения, аудио, видео
Transformer-модели	Используют механизм внимания и могут выступать основой разных генеративных систем	Текст, код, изображения, видео

Авторегрессионная генерация

Авторегрессионная модель формирует последовательность поэлементно. На каждом шаге она оценивает вероятности возможных следующих элементов с учётом уже имеющегося контекста.

Для текста элементом обычно является токен. Пусть модель получила фразу «Искусственный интеллект используется для». Она оценивает вероятности возможных продолжений и выбирает один из токенов, после чего повторяет процесс уже с расширенным контекстом.

Выбор следующего элемента может быть детерминированным или вероятностным. Изменяя параметры генерации, можно получить более предсказуемые или более разнообразные результаты.

Автокодировщики

Автокодировщик состоит из двух основных частей: энкодера и декодера. Энкодер преобразует входные данные в компактное скрытое представление, а декодер восстанавливает исходные данные из этого представления.

Компонент	**Назначение**
Encoder	Преобразует вход в скрытый вектор
Latent space	Компактное представление существенных свойств данных
Decoder	Восстанавливает объект из скрытого представления

Обычный автокодировщик прежде всего решает задачу реконструкции. Для генерации новых объектов необходимо, чтобы скрытое пространство было организовано так, чтобы из него можно было корректно выбирать новые точки.

Вариационные автокодировщики VAE

Variational Autoencoder (VAE) — генеративная модификация автокодировщика. Вместо фиксированной точки энкодер описывает вероятностное распределение в скрытом пространстве. При обучении модель одновременно стремится хорошо реконструировать данные и поддерживать упорядоченную структуру скрытого пространства.

После обучения из скрытого распределения можно выбрать новую точку и передать её декодеру, получив новый объект. В загруженных учебных материалах отдельно предусмотрено изучение автокодировщиков, VAE и CVAE как современных нейросетевых подходов.

Условные генеративные модели

Условная генеративная модель создаёт данные не полностью свободно, а с учётом дополнительного условия. Условием может быть класс, текстовое описание, изображение, стиль, категория или иной контекст.

Например, conditional VAE может генерировать объект заданного класса, а модель text-to-image создаёт изображение по текстовому описанию. Управляемость генерации является одним из ключевых свойств современных генеративных систем.

Генеративно-состязательные сети GAN

Generative Adversarial Network (GAN) включает две нейронные сети: генератор и дискриминатор. Генератор создаёт синтетические примеры, а дискриминатор пытается отличить реальные данные от созданных генератором.

Упрощённая схема GAN

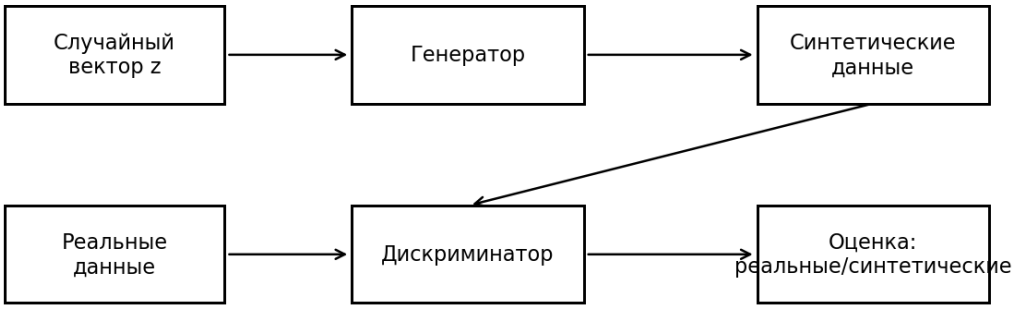


Рисунок 2 – Упрощённая схема генеративно-состязательной сети

В ходе обучения генератор стремится создавать всё более правдоподобные примеры, а дискриминатор — лучше различать реальные и синтетические данные. Такой состязательный процесс может приводить к высокому качеству генерации, особенно для изображений.

Преимущества GAN	**Ограничения GAN**
Высокое визуальное качество во многих задачах	Обучение может быть нестабильным

Быстрая генерация после обучения	Возможен mode collapse — недостаточное разнообразие
Гибкость для условной генерации	Сложнее объективно оценивать качество
Применение к изображению и другим данным	Требует тщательной настройки архитектуры и обучения

Диффузионные модели

Диффузионные модели стали важным классом генеративных моделей для изображений, аудио и видео. Идея основана на последовательном зашумлении данных и обучении модели обратному процессу — постепенному восстановлению структуры из шума.

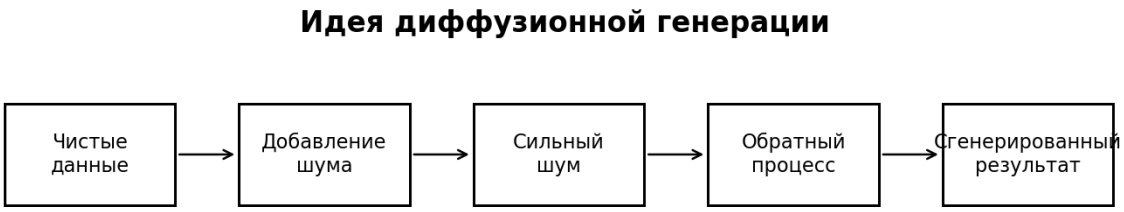


Рисунок 3 – Концептуальная схема диффузионной генерации

Во время генерации процесс начинается с случайного шума. Модель шаг за шагом удаляет шум, направляя результат к объекту, согласованному с условием. При text-to-image таким условием является текстовое описание.

В загруженной англоязычной хрестоматии приведён пример современной модели генерации видео, где видео сначала переводится в компактное скрытое пространство, представляется пространственно-временными патчами, а генерация выполняется диффузионным трансформером.

GAN, VAE и диффузионные модели: сравнение

Критерий	**VAE**	**GAN**	**Диффузионные модели**
Принцип	Вероятностное скрытое пространство	Состязательное обучение	Обратное восстановление из шума
Качество реконструкции/генерации	Стабильное, иногда более сглаженное	Может быть очень реалистичным	Обычно высокое
Стабильность обучения	Относительно высокая	Может быть сложной	Обычно стабильнее GAN
Скорость генерации	Быстрая	Быстрая	Может требовать нескольких шагов
Скрытое пространство	Явно используется	Не обязательно удобно интерпретируется	Часто работает в пиксельном или латентном пространстве

Генерация текста

Генерация текста основана на моделировании последовательностей естественного языка. Современные языковые модели часто используют архитектуру Transformer и обучаются

предсказывать токены или восстанавливать части текста в зависимости от постановки обучения.

Генерируемый текст может использоваться для черновиков, объяснений, суммаризации, перевода, преобразования стиля, структурирования информации и программной помощи. Однако правдоподобность текста не означает его фактическую правильность.

Например, модель может сформировать связное объяснение научного понятия, но допустить ошибку в числах, ссылках или фактах. Поэтому результат необходимо проверять по надёжным источникам.

Генерация программного кода

Программный код можно рассматривать как структурированную последовательность токенов. Генеративная модель способна создавать функции, объяснять код, предлагать тесты и преобразовывать программу между языками.

При этом сгенерированный код необходимо проверять: возможны логические ошибки, неправильное использование библиотек, уязвимости или несоответствие требованиям. Генерация кода должна дополняться тестированием, анализом безопасности и контролем разработчика.

Генерация изображений

Генерация изображений может выполняться GAN, VAE и особенно диффузионными моделями. Пользователь задаёт текстовое условие, а модель формирует новое визуальное содержание.

Современные системы часто работают не непосредственно с исходными пикселями на всех этапах, а используют компактное латентное пространство. Это уменьшает вычислительные затраты и позволяет эффективнее обучать и применять модель.

Режим	**Описание**
Text-to-image	Изображение создаётся по текстовому описанию
Image-to-image	Исходное изображение преобразуется с сохранением части структуры
Inpainting	Заполняется выбранная область изображения
Outpainting	Расширяется изображение за исходные границы
Conditional generation	Результат контролируется дополнительными условиями

Генерация аудио

Генеративные модели могут создавать речь, музыку и другие звуковые сигналы. Задача может быть условной: текст преобразуется в речь, голосовой сигнал модифицируется, а музыкальный фрагмент создаётся по описанию или заданной структуре.

Аудиогенерация требует особого внимания к правам на голос, персональным данным и возможности введения слушателя в заблуждение. Синтетический голос должен использоваться с разрешением и в соответствии с правовыми и этическими требованиями.

Генерация видео

Генерация видео сложнее генерации отдельного изображения, поскольку необходимо поддерживать согласованность не только внутри кадра, но и во времени. Модель должна учитывать движение объектов, изменение камеры, освещения и причинно-временные зависимости между кадрами.

В англоязычной хрестоматии рассмотрен подход, в котором видео сжимается в латентное пространство, делится на пространственно-временные патчи, а диффузионный трансформер обучается восстанавливать чистое представление из зашумленного. Такой подход демонстрирует объединение идей компрессии, Transformer и diffusion.

Мультимодальный генеративный ИИ

Мультимодальная модель работает с несколькими типами данных. Например, она может получать текст и изображение, анализировать их совместно и формировать текстовый ответ; либо получать текст и создавать изображение или видео.

Вход	**Выход**	**Пример задачи**
Текст	Текст	Суммаризация документа
Текст	Изображение	Создание иллюстрации
Изображение	Текст	Описание изображения
Текст + изображение	Текст	Анализ схемы или документа
Текст	Аудио	Синтез речи
Текст / изображение	Видео	Генерация видеосцены

Prompt как условие генерации

Prompt — инструкция или контекст, передаваемый модели для управления результатом. Формулировка запроса влияет на содержание, формат и уровень детализации ответа.

Для текстовой модели полезно явно задавать задачу, входные данные, ограничения и ожидаемый формат. Для генерации изображений описание может включать объект, композицию, освещение, стиль и другие параметры. Более системно методы построения запросов рассматриваются в отдельной теме Prompt Engineering.

Параметры разнообразия генерации

Генеративная модель часто формирует вероятностное распределение возможных результатов. Управление выбором из этого распределения позволяет менять степень вариативности.

Режим	**Результат**
Более детерминированный выбор	Ответы стабильнее и предсказуемее
Более вероятностный выбор	Результаты разнообразнее, но риск отклонений выше
Фиксация seed в некоторых генераторах	Помогает воспроизводить близкий результат при одинаковых условиях

Параметры генерации нельзя оценивать как «лучше» или «хуже» вне задачи. Для формального технического текста обычно требуется большая стабильность, а для творческого поиска может быть полезно разнообразие.

Оценка качества генеративного результата

Оценка генеративных моделей сложнее, чем оценка обычного классификатора, поскольку часто не существует единственного правильного результата. Необходимо учитывать несколько критериев одновременно.

Критерий	**Вопрос**
Соответствие запросу	Выполнено ли условие пользователя?

Корректность	Нет ли фактических или логических ошибок?
Качество формы	Грамотен ли текст, визуально ли целостно изображение?
Разнообразие	Не повторяет ли модель один и тот же шаблон?
Безопасность	Нет ли недопустимого, вредного или конфиденциального содержания?
Происхождение и права	Можно ли использовать результат в выбранном контексте?

Галлюцинации и фактическая достоверность

Для текстовых генеративных моделей характерна возможность сформировать правдоподобное, но фактически неверное утверждение. Такое явление часто называют галлюцинацией модели.

Причина заключается в том, что модель оптимизируется для построения вероятного продолжения или ответа в рамках обученного представления, а не гарантированного обращения к проверенной базе фактов.

Поэтому фактические сведения, числа, ссылки, нормативные положения и научные утверждения необходимо проверять по первичным или надёжным источникам.

Смещения и качество обучающих данных

Генеративная модель наследует свойства данных, на которых она обучалась. Если данные содержат ошибки, дисбаланс, устаревшие сведения или систематические искажения, это может проявляться в результатах.

Ответственное применение требует учитывать происхождение данных, область применимости модели и возможные различия качества для разных языков, доменов и категорий пользователей.

Авторское право и происхождение контента

Генеративный ИИ поднимает вопросы прав на обучающие материалы, сгенерированный контент и использование результатов. Практическая политика зависит от законодательства, лицензий и условий конкретного сервиса.

В учебной и научной работе необходимо прозрачно указывать использование генеративных инструментов, если этого требуют правила организации или издания, и не представлять автоматически созданный материал как проверенный первичный источник.

Персональные и конфиденциальные данные

Нельзя без необходимости передавать внешнему генеративному сервису персональные данные, пароли, ключи доступа, внутренние документы, закрытый исходный код и другую конфиденциальную информацию.

Если генеративный ИИ используется в организации, необходимо определить разрешённые типы данных, права доступа, правила хранения запросов и результатов, а также процедуры проверки и публикации.

Генеративный ИИ и информационная безопасность

Положительное применение	**Риск / ограничение**
Подготовка черновиков аналитических отчётов	Необходима проверка фактов
Суммаризация журналов и описаний инцидентов	Нельзя передавать секретные логи во внешний сервис без разрешения

Объяснение программного кода	Сгенерированный код может содержать ошибки и уязвимости
Создание обучающих материалов	Нужно проверять корректность и происхождение контента
Помощь аналитикам SOC в структурировании текста	Модель не заменяет техническую верификацию инцидента

Риски злоупотребления генеративными технологиями

Генеративные модели могут использоваться для создания вводящего в заблуждение контента, масштабирования социальной инженерии и подделки медиаматериалов. Поэтому средства генерации должны применяться с учётом безопасности и ответственности.

Защитные меры включают обучение пользователей, проверку источников, многофакторную аутентификацию, контроль публикаций и использование технических средств проверки контекста и происхождения данных. Нельзя полагаться только на визуальную или стилистическую правдоподобность материала.

Генеративный ИИ в образовании и науке

Генеративные модели могут помогать объяснять материал, создавать черновые примеры, формировать вопросы, структурировать идеи и поддерживать программирование. Однако они не должны подменять понимание предмета и самостоятельную проверку.

В научной работе модель может помочь отредактировать формулировку, структурировать текст или предложить код анализа, но научные выводы, ссылки, экспериментальные результаты и интерпретация должны подтверждаться исследователем.

Генеративный ИИ в работе с документами

Практический сценарий: пользователь имеет длинный документ и хочет получить краткое содержание. Безопасная схема включает передачу только разрешённого содержания, формулирование задачи, получение чернового резюме, сравнение с исходным документом и ручное исправление.



Рисунок 4 – Рекомендуемый рабочий процесс при использовании генеративного ИИ

Такой подход подчёркивает, что модель является инструментом подготовки результата, а ответственность за итоговое решение остаётся у человека или организации, применяющей систему.

Пример сравнения генеративных подходов

Задача	**Подход**	**Почему подходит**
Создание вариаций изображения	VAE / GAN / diffusion	Моделируют распределение визуальных данных
Генерация реалистичных изображений по тексту	Диффузионная условная модель	Высокое качество и управляемость условием

Генерация текста и кода	Авторегрессионный Transformer	Последовательная генерация токенов
Компактное скрытое представление и генерация	VAE	Явное вероятностное латентное пространство
Синтез видео	Диффузионный Transformer и другие генеративные архитектуры	Учитывает пространственно-временную структуру

Ограничения генеративных моделей

– Результат может быть фактически неверным, хотя выглядит убедительно.
– Качество зависит от обучающих данных и области применения модели.
– Сложно гарантировать полное отсутствие смещений и нежелательных результатов.
– Некоторые модели требуют значительных вычислительных ресурсов.
– Оценка качества часто включает человеческое экспертное суждение.
– Возникают вопросы авторского права, конфиденциальности и ответственности.
– Модель не должна автоматически получать доступ к критичным действиям без контроля и ограничений.

Практическая схема выбора генеративного инструмента

Вопрос	**Что определить**
Какой тип данных нужен?	Текст, изображение, код, аудио или видео
Нужна ли генерация или анализ?	Генеративная модель не всегда нужна для классификации
Есть ли конфиденциальные данные?	Выбрать разрешённую среду и режим работы
Нужен ли строгий фактологический ответ?	Предусмотреть проверку по источникам
Нужно ли повторять результат?	Контролировать параметры и воспроизводимость
Какие ограничения использования?	Лицензии, политика организации, авторские права

Основные выводы

– Генеративный искусственный интеллект создаёт новый контент на основе закономерностей обучающих данных.
– К основным генеративным подходам относятся авторегрессионные модели, VAE, GAN и диффузионные модели.
– Современные системы могут генерировать текст, изображения, код, аудио и видео.
– Условная генерация позволяет управлять результатом с помощью текста, класса, изображения или другого контекста.
– Качество генерации необходимо оценивать по соответствию задаче, корректности, разнообразию и безопасности.
– Правдоподобный ответ генеративной модели может быть неверным и требует проверки.
– При работе с генеративным ИИ необходимо учитывать конфиденциальность, авторское право, безопасность и ответственность.

Контрольные вопросы

1. Что такое генеративный искусственный интеллект и чем генеративная модель отличается от дискриминативной?

2. Каковы основные принципы авторегрессионной генерации, VAE, GAN и диффузионных моделей?

3. В чём различие между обычным автокодировщиком и вариационным автокодировщиком?

4. Как взаимодействуют генератор и дискриминатор в GAN и какие проблемы могут возникать при обучении?

5. Как устроен общий принцип диффузионной генерации и почему он применяется для изображений и видео?

6. Какие особенности имеют генерация текста, изображений, аудио, видео и программного кода?

7. Что такое условная и мультимодальная генерация и какую роль в них играет prompt?

8. По каким критериям следует оценивать качество генеративного результата?

9. Какие риски связаны с галлюцинациями, конфиденциальными данными, авторским правом и вводящим в заблуждение контентом?

10. Как организовать безопасный и ответственный рабочий процесс применения генеративного ИИ в учебной, научной или профессиональной задаче?

Список источников

1. Учебное пособие «Основы искусственного интеллекта». Глава 11 «Генеративный искусственный интеллект»: Generative AI, генерация текста, изображений, аудио и видео.

2. Беликова Е.К., Гудилина И.С., Круглова М.А., Саратовская Л.Б., Ямпольская А.В. English Reader in Artificial Intelligence = Хрестоматия по искусственному интеллекту на английском языке. Учебное пособие. Москва, 2024. – 171 с.

3. Бессмертный И.А., Авдюшина А.Е., Кугаевских А.В., Королева Ю.А., Рущенко Н.Г. Системы искусственного интеллекта. Методическое пособие. Санкт-Петербург: Университет ИТМО, 2024. – 281 с.

4. Андреянов Н.В., Евдокимова Т.С., Павлов А.Д., и др. Методы искусственного интеллекта. Учебное пособие. Казань: КНИТУ-КАИ, 2024. – 392 с.