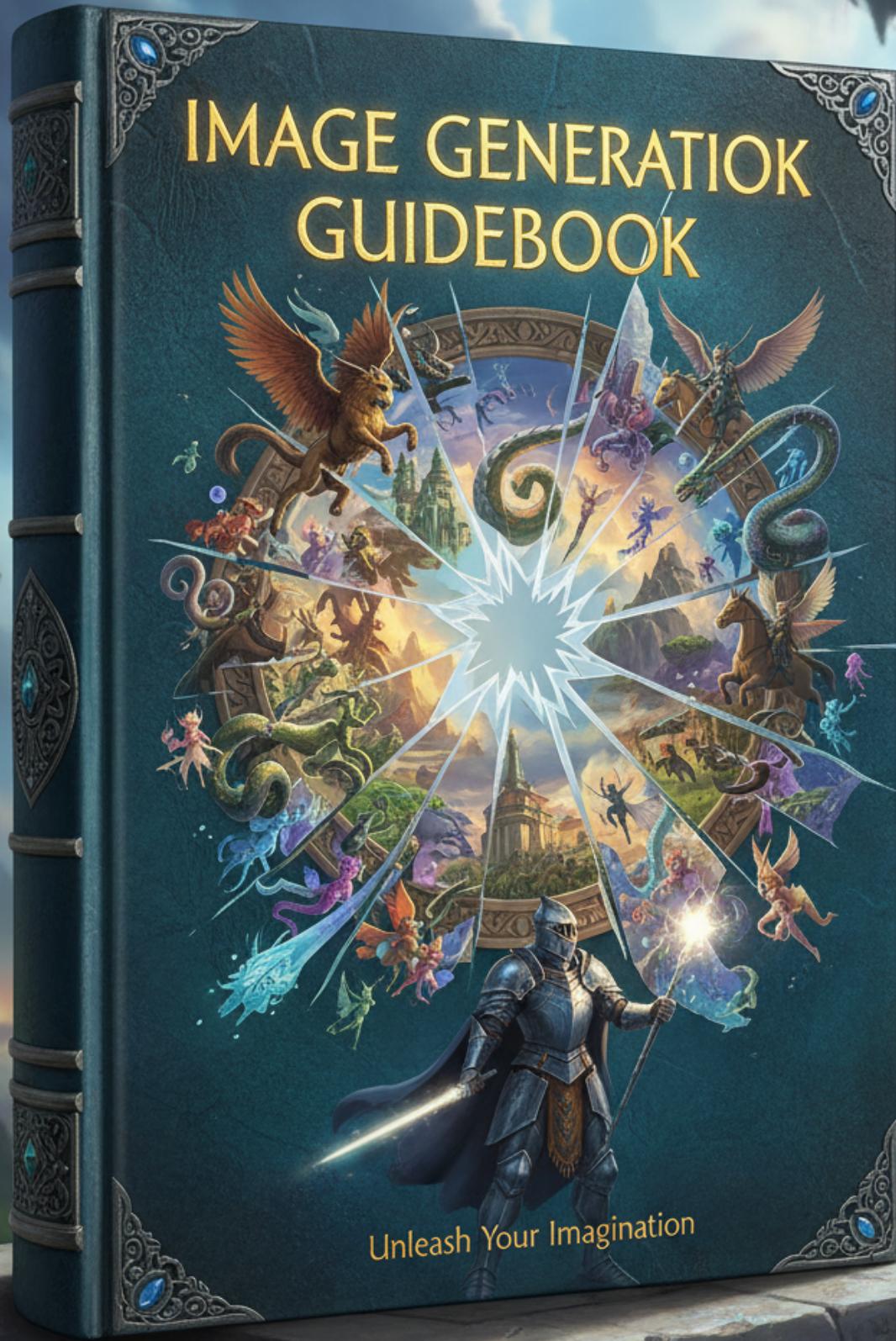


IMAGE GENERATION GUIDEBOOK

Unleash Your Imagination



Введение

Генерация изображений с помощью нейросетей перестала быть инструментом только для дизайнёров и технарей — сегодня это мощный и удобный способ создавать визуальный контент любого уровня: от скетчей и концептов до реалистичных фото, 3D-персонажей и рекламных визуалов. За последние пару лет экосистема моделей выросла в полноценный мир, где каждая нейросеть имеет свой характер, стиль и область применения.

В этом гайде мы в самом начале разберём историю развития отрасли генерации изображений, быстро пробежимся по технологиям, на основе которых строится происходит генерация и далее в рамках постепенного обзора перейдем к самым популярным и актуальным инструментам для создания изображений на данный момент.

Ты узнаешь, чем они отличаются, какой результат дают, как ими пользоваться и как выбирать модель под конкретную задачу — будь то реалистичные фото, стилизованная графика, анимация или грамотная работа с художественным направлением.

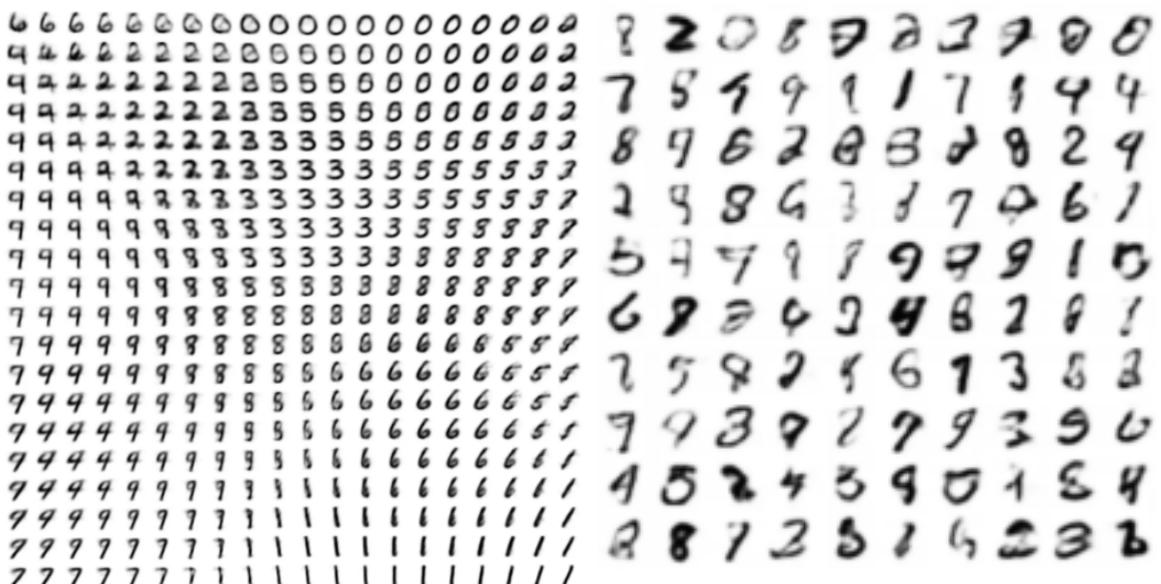
Гайд построен так, чтобы шаг за шагом провести тебя через основы и показать практические приёмы, которые действительно работают. Мы поговорим о структуре промтov, параметрах качества, пропорциях, нюансах света, о том, как управлять стилем, и как добиться предсказуемого результата, используя силу современных моделей.

Этот материал подойдёт всем: от тех, кто впервые открывает генерацию, до продвинутых пользователей, которые хотят глубже понимать возможности каждой модели и собирать предсказуемые, красивые и профессиональные результаты.

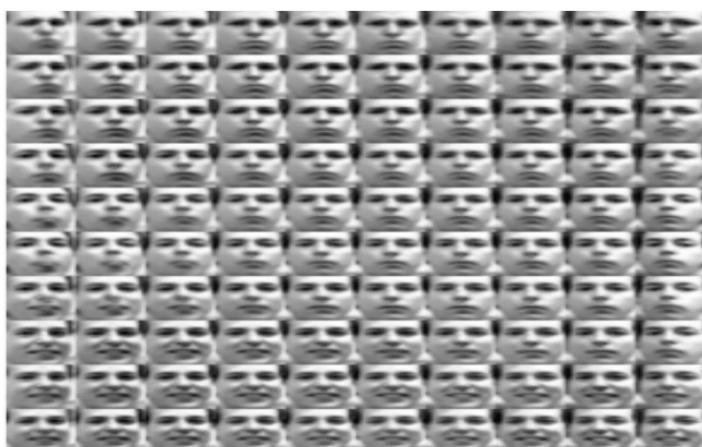
История и база в генерации

Первые попытки начать генерировать изображения начались примерно с того же момента как появились первые мат модели персептрана для создания первых ИИ в 1960-е года, но вычислительные мощности и большое количество контекстуальной информации в изображениях просто не позволяло полноценно работать с ними даже на уровне распознавания текстов или образов.

Первые более-менее решенные задачи связанны были с распознаванием и генерацией рукописных черно-белых цифр и лиц в 2010-е, но и они были далеки от идеала. Требовалось создать скрытое пространство, в котором где для каждого класса (например изображение цифры) было бы свое распределение образов, из которых модель могла бы брать точку и превращать в нужную цифру. Сложность в том, что если мы будем брать точки между распределениями, то будем получать невнятное представление случайных пикселей, а если брать точки напрямую из распределений, то будем получать те же изображения цифр, которые были в обучающей выборке.



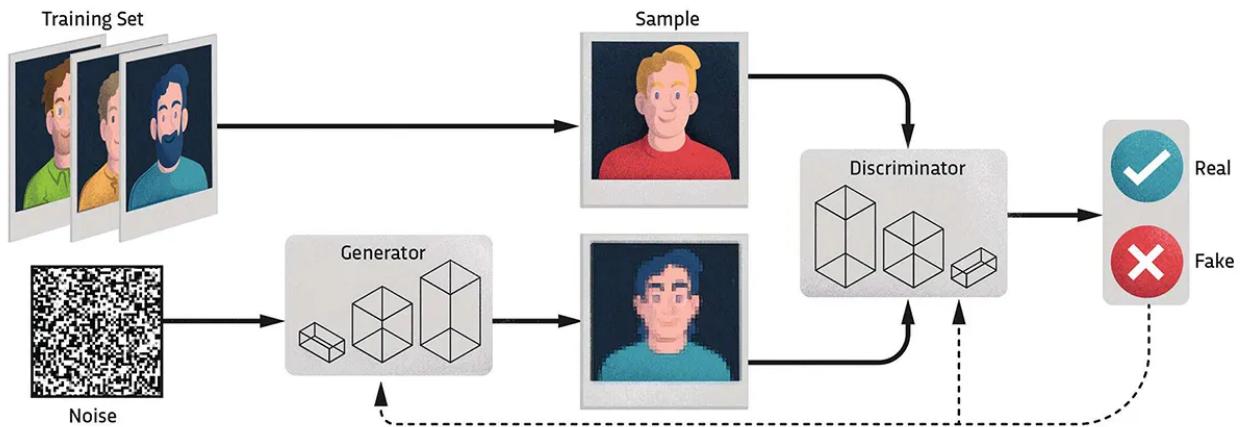
Но в 2014 года вышла статья Generative Adversarial Networks (GAN) которая показала первый прорыв в области, в ней был представлен подход вызволяющий генерировать уникальные черно-белые лица людей в низком разрешении.



База

Концептуально на сегодняшний день существует 2 основных подхода в генерации на основе которых сейчас держатся все современные генеративные модели

1. Давно изобретенный, но сильно переработанный GAN



В основе GAN лежат две нейронные сети, которые учатся, постоянно соревнуясь друг с другом:
Генератор (Generator): "Фальшивомонетчик". Его задача — создавать поддельные изображения из случайного шума.

Дискриминатор (Discriminator): "Следователь". Его задача — отличать настоящие изображения (из обучающей выборки) от поддельных (созданных генератором).

Процесс обучения представляет собой итеративную "гонку вооружений":

Шаг 1: Обучение Дискриминатора

Генератор создает партию фальшивых изображений.

Дискриминатору подается смесь реальных и фальшивых изображений.

Дискриминатор пытается их классифицировать, и после каждой попытки он получает "обратную связь":

"Вот это изображение было настоящим, а ты сказал, что фальшивое — ошибся!"

"Вот это было фальшивкой, а ты принял за настоящую — учись!"

На основе этих ошибок веса дискриминатора немного корректируются, и он становится немного лучше в распознавании подделок.

Шаг 2: Обучение Генератора

Теперь мы "замораживаем" дискриминатор. Он становится таким экспертом с постоянным уровнем навыков.

Генератор создает новую партию фальшивых изображений и пытается "обмануть" замороженного дискриминатора.

Дискриминатор оценивает эти изображения.

Ошибка дискриминатора (его неспособность распознать подделку) передается генератору как сигнал для обучения:

"Вот эти фальшивки дискриминатор с легкостью раскрыл — значит, они плохие, не делай так."

"А вот эти фальшивки прошли проверку — отлично, развивай это направление!"

На основе этого сигнала веса генератора корректируются, и он учится создавать более правдоподобные изображения.

Цикл и равновесие

Эти два шага повторяются десятки или сотни тысяч раз.

Сначала генератор создает просто шум, а дискриминатор легко его разоблачает.

Постепенно генератор учится угадывать простые структуры (например, общие формы).

Дискриминатор, в ответ, становится более внимательным к деталям.

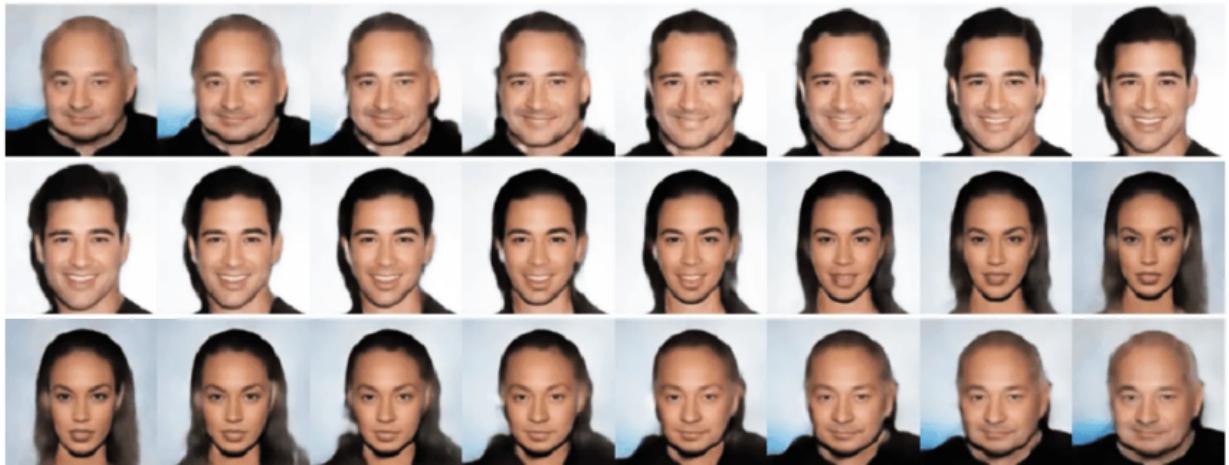
В идеале, процесс достигает состояния равновесия (Nash Equilibrium), когда генератор создает изображения настолько качественные, что дискриминатор не может отличить их от настоящих, и просто угадывает с вероятностью 50%.

Наглядная аналогия

Представьте, что:

Генератор — студент-художник, подделывающий картины Ван Гога.

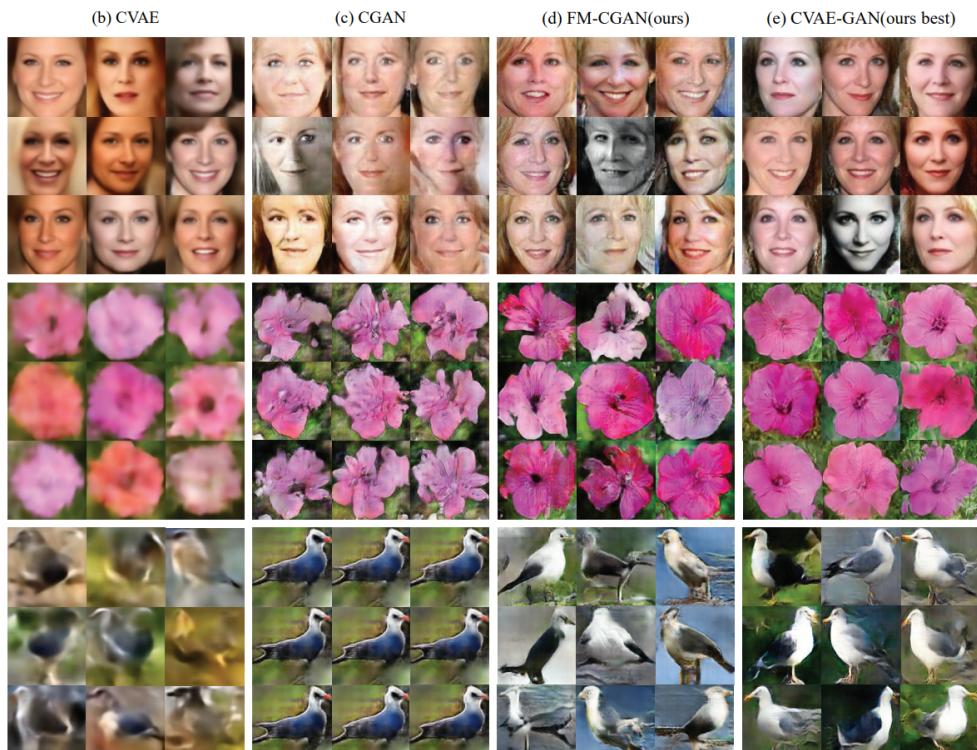
Дискриминатор — искусствовед, эксперт по Ван Гогу.



Сначала студент рисует плохо, и эксперт сразу видит подделку. Но с каждой неудачей студент получает фидбэк ("слишком яркие цвета", "не та фактура мазка") и улучшает свою технику. В то же время, эксперт, видя все более качественные подделки, вынужден сам становиться более проницательным и изучать мельчайшие детали. В итоге, студент начинает писать картины, которые практически неотличимы от оригиналов для самого эксперта.

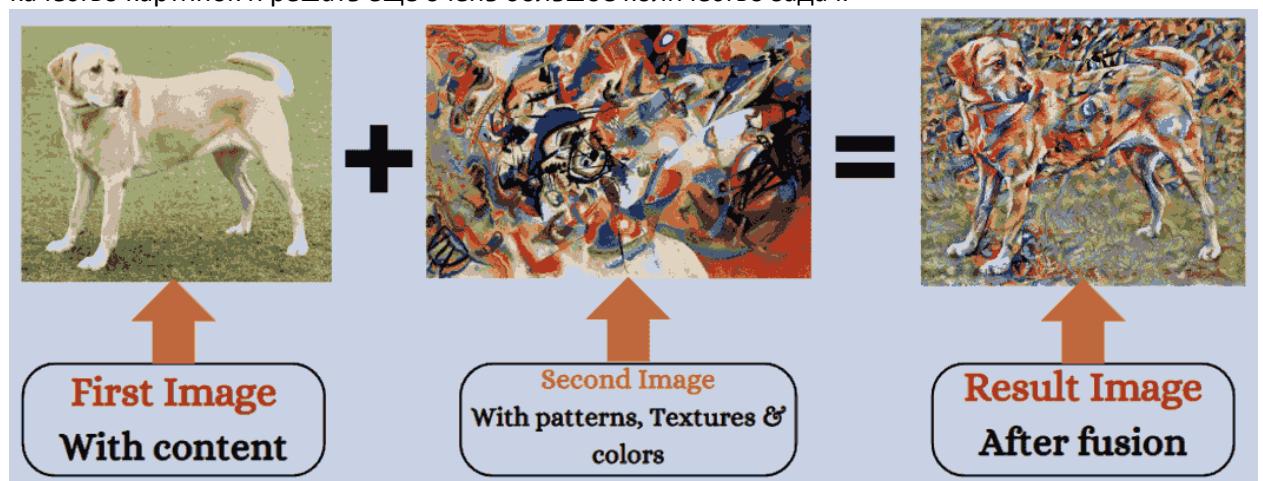
Главная сложность в обучении GAN — поддерживать это хрупкое равновесие. Если одна из сетей станет слишком сильной (например, дискриминатор будет всегда "выигрывать"), обучение второй сети остановится. Эта нестабильность — одна из ключевых причин, почему на смену GAN во многих задачах пришли более стабильные диффузионные модели.

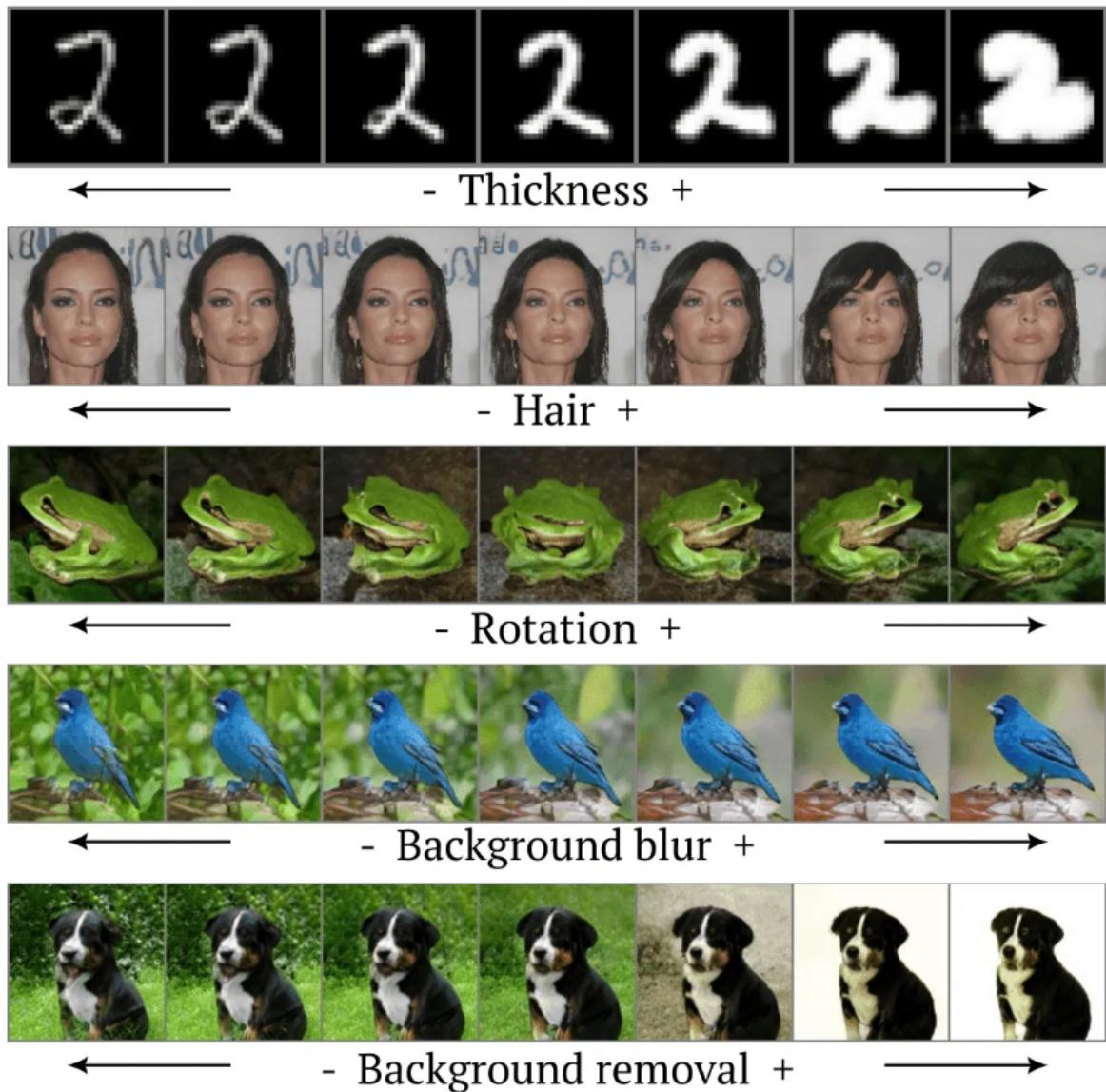
У такого подхода очень много минусов — например генератор может научиться генерировать одно идеальное изображение, которое дискриминатор каждый раз мог считать за реальные и таких проблем там было очень много, за 6 лет работы с GAN прогресс шел очень медленно относительно прогресса в наше время, но был представлен результат с лицами несуществующих людей <https://thispersonnotexist.org/> который пробудил интерес бизнеса к сфере генерации изображений.



Четырехлетний прогресс GAN подхода

Такой подход позволял переносить стилистику с картин, по маскам рисовать пейзажи, улучшать качество картинок и решать еще очень большое количество задач.





Однако в силу сложности обучения и малой вариативности GAN моделей начали искать альтернативные подходы. И нашли..

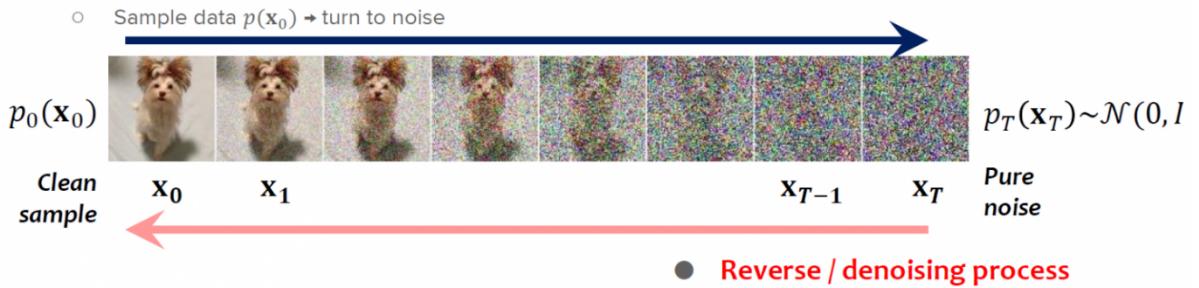
2. В 2020 году был представлена статья и вместе с ней и подход «Denoising Diffusion Probabilistic Models» (DDPM).

Диффузионный подход — это принципиально иной, но не менее гениальный механизм, который сейчас находится на пике популярности. Если GAN — это «противостояние», то диффузионные модели — это «очищение от шума».

Основная идея: От шума к порядку через постепенное очищение

Denoising diffusion models

● Forward / noising process



- Sample noise $p_T(x_T) \rightarrow$ turn into data

Принцип работы диффузионных моделей можно разделить на два ключевых процесса: прямой (forward process) и обратный (reverse process).

Прямой процесс (Forward Process): "Испортить изображение"

Представьте, что у вас есть четкая фотография.

Вы берете это чистое изображение и добавляете к нему небольшое количество случайного шума (как будто слегка засыпаете его песком).

Вы повторяете этот шаг много раз (например, 1000 раз), каждый раз добавляя еще немного шума. В конце концов, исходное изображение полностью превращается в абсолютно случайный шум, неотличимый от статического помеха на телевизоре.

Этот процесс детерминированный и не требующий обучения. Нейросеть здесь не учится, мы просто пошагово разрушаем данные по заранее известной формуле.

Зачем это нужно? Мы показываем модели, как выглядит изображение на каждой стадии его "зашумления".

Обратный процесс (Reverse Process): "Восстановить изображение" — вот где происходит магия!

Это и есть этап обучения модели.

Мы берем совершенно случайный шум (результат прямого процесса).

Задача нейросети — предсказать, как выглядел этот шум на один шаг назад, то есть предугадать, какой шум нужно убрать, чтобы картинка стала немного четче.

Модель делает это предсказание, и мы сравниваем его с тем, какой шум мы на самом деле добавляли на соответствующем шаге прямого процесса.

Разница между предсказанным и реальным шумом — это ошибка (loss), на основе которой и обучается нейросеть. По сути, мы учим ее "откатывать" процесс зашумления.

Проще говоря, в процессе обучения модель отвечает на вопрос: "Если я вижу вот эту зашумленную картинку, то какой именно шум мне нужно убрать, чтобы она стала немного чище?"

Как происходит генерация нового изображения?

После обучения модель готова к работе:

Вы начинаете с абсолютно случайного шума.

Модель, шаг за шагом, последовательно удаляет из этого шума предсказанное количество "песка".

На первом шаге из хаоса начинают проступать лишь общие цвета и формы.

С каждым следующим шагом детали становятся все четче и осмысленнее.

После выполнения всех шагов (например, тех же 1000) изначальный шум превращается в чистое, детализированное изображение.

Наглядная аналогия

Представьте мраморную глыбу, в которой спрятана статуя.

Прямой процесс — это постепенное, послойное покрытие статуи густой грязью, пока она не превратится в бесформенный ком.

Обучение — это наблюдение за тем, как мастер-скульптор очищает этот ком. Вы запоминаете, как он на каждом этапе снимает ровно тот слой грязи, который скрывает форму. Генерация — это когда вы, научившись, берете новую бесформенную глыбу (шум) и сами, шаг за шагом, "откалываете" от нее лишнее, следуя выученной технике, и в итоге получаете новую статую.

Именно на этом принципе работают все современные публично представленные модели. Они взяли базовый диффузионный процесс и добавили к нему ключевой компонент — текстовые промты, которые направляют процесс очищения от шума в нужном смысловом направлении.

От теории к практике

Теперь понимая принцип работы генеративных моделей, перейдем к практикам и гайдам по их использованию. В этом разделе я в начале расскажу про использование открытых моделей, которые можно локально запускать у себя на ПК без интернета, инструменты, которые можно применять вместе с этими моделями, а также сервисы для их использования. Далее расскажу то же, но про платные сервисы и лучшие практики для их использования.

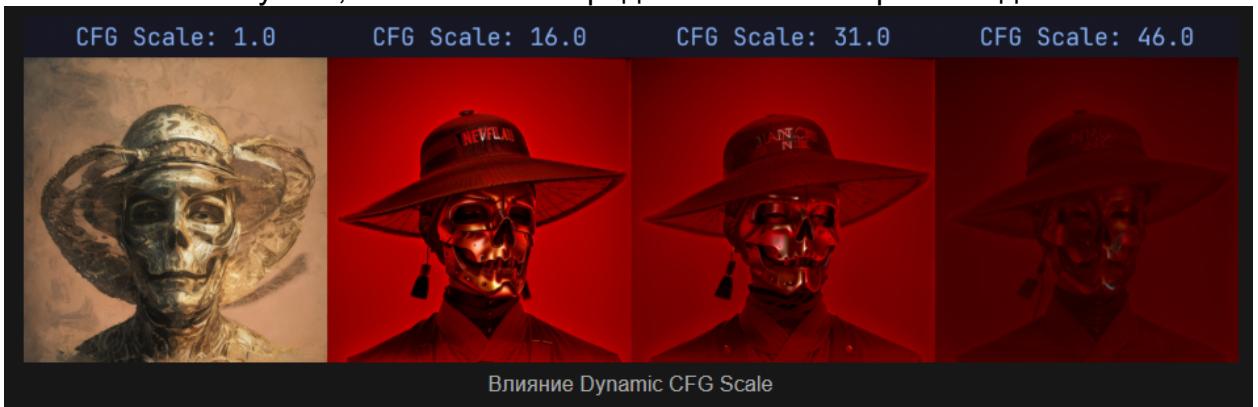
Основные параметры генерации

В современных сервисах редко модно встретить что-то кроме поля для промта НО иногда разработчики дают доступ и к другим полям генерации, далее я опишу все, в том числе те что под капотом моделей.

Prompt — основная информация при генерации то что должно находиться на изображении. Не рекомендуется писать слишком длинные промты, далее я расскажу как лучше оформлять промты

Negative prompt — то чего НЕ ДОЛЖНО быть на изображении

CFGSCALE— мера того на сколько сильно промт действует на генерацию на уровне математики расшумления. Слишком низкие значения будут делать картинку серой и не интересной, слишком высокие высвечивать цвета либо затемнять. Для каждой модели этот параметр свой и не рекомендуется менять его более чем на 2 пункта, относительно предложенного авторами модели.



STEPS — Количество шагов расшумления. Чем больше шагом тем детальнее картинка, но дольше время генерации.

SEED - Этот параметр определяет случайный шум на основании которого генерируется картинка. Само значение параметра ни на что не влияет НО при одинаковом промте и значении seed будет получаться одна и та же картинка каждый раз, если хотябы чуть-чуть поменять seed то изображение уже будет другим. Так же можно фиксировать seed одним значением и менять промт, это будет вносить лишь минимальные изменения в картинку (но такой подход не всегда работает стабильно).

clipSkip – Параметр важный для обучения собственных моделей, то на сколько скрытые представления влияют на промт во время генерации. Для реалистичных фото принято ставить 1, для мультишных 2.

sampler - Определяет расписание добавления шума во время генерации. Сугубо технический параметр, рекомендуется ставить по умолчанию **EULER A**. Либо тот что рекомендует автор модели.

Основная логика написания промта

Подумайте о том, какие детали имеют значение. Всё, что вы упустите, будет рандомизировано.

Основной объект: Мужчина, женщина, кот и т. д.

Второстепенные объекты + ландшафты, окружение: На холме, в больнице, посреди океана и т. д.

Стиль референс: Графический стиль, аниме, фотореалистичный стиль и т. д.

Материал: фотография, живопись, иллюстрация, скульптура, каракули, гобелен и т. д.

Композиция: портрет, крупный план, вид с высоты птичьего полета и т. д. Тема: человек, животное, персонаж, локация, предмет и т. д.

Окружающая среда: в помещении, на улице, на Луне, в Нарнии, под водой, в Изумрудном городе и т. д.

Освещение: мягкое, рассеянное, пасмурное, неоновое, студийное и т. д.

Цвет: насыщенный, приглушенный, яркий, однотонный, красочный, чёрно-белый, пастельный и т. д.

Настроение: Уравновешенное, спокойное, хриплое, энергичное и т. д.

ЧТО?	ФОТО	КАКОЕ ФОТО?	В ДВИЖЕНИИ
КТО?	МОДЕЛЬ	КАК ВЫГЛЯДИТ?	МОДЕЛЬ-АЛЬБИНОС С БОЛЬШИМИ СЕРЫМИ ГЛАЗАМИ, ДЛИННЫМИ ПРЯмыми БЕЛЫМИ ВОЛОСАМИ, СЕРО-БЕЛОЙ КОЖЕЙ
ЧТО ДЕЛАЕТ?	ТАНЦУЕТ	КАК?	СТРАННО
ГДЕ?	В ЦЕРКВИ	КАК ВЫГЛЯДИТ?	В БЕТОННОЙ КАТОЛИЧЕСКОЙ ЦЕРКВИ
В ЧЁМ?	В ПЛАТЬЕ	В КАКОМ?	БЕЛОЕ ШИФОНОВОЕ ПЛАТЬЕ СО МНОЖЕСТВОМ СЛОЁВ ПОЛУПРОЗРАЧНОЙ ТКАНИ
СВЕТ?	СУМЕРКИ	ЦВЕТ?	ГЛУБОКИЙ СИНИЙ, СЕРЫЙ И ЧЁРНЫЙ ЦВЕТ
НАСТРОЕНИЕ?	ДЕПРЕССИВНОЕ	ЭСТЕТИКА	DREAMCORE



Opensource модели.

Все модели этой группы можно найти на <https://civitai.com/> с примерами генерации и различными чекпойнтами от людей которые доделывали свои версии на основе нижеописанных архитектур

SD 1.4 / 1.5

Минимально достаточный уровень для знакомства с генеративными моделями — это Stable Diffusion 1.4 и 1.5. Они небольшие по размеру, просты в установке и дообучении, и их можно запустить даже на слабом ПК или процессоре (правда, медленно).

Однако у этих моделей есть существенные ограничения:

- Они часто ошибаются в анатомии — руки с лишними пальцами, ноги разной длины, «плавающие» глаза и другие артефакты встречаются регулярно;
- Качество резко падает при разрешении выше 768×768 пикселей: композиция разваливается, детали теряются, появляются шумы и искажения;
- Поскольку обучались на нефильтрованной выборке из интернета, они могут генерировать крайне неприемлемый, вредоносный или травмирующий контент — особенно если промпт написан расплывчато или провокационно.

По сути, SD 1.4/1.5 — это «учебный полигон»: отлично подходит для понимания принципов работы, но требует осторожности, ручной доработки и критического взгляда на результат.

Практики написании промтта таки старый моделей крайне специфичны. Изображение должно быть написано рядом словосочетаний либо отдельных слов через запятую с прямым указанием важности слов путем нескольких выделений скобками, а также коэффициентом веса важности слова. Например для изображения ниже созданного на sd 1.5



Prompt - (best quality:1.4), anime, solo, large breasts, thick thighs, brown eyes, sitting, ankle bracelet, waterfall, river, rocks, mountains, ((sunbeam)), flower crown, beautiful detailed face

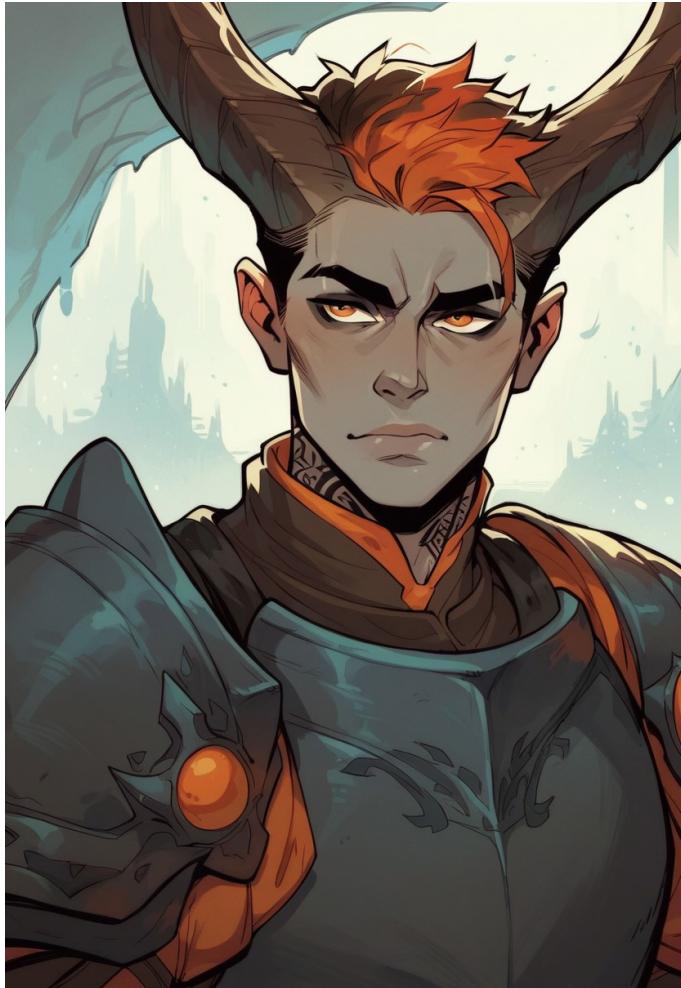
Negative prompt - (worst quality, low quality:1.4), multiple views, (((watermark, signature))), floating hair, easynegative

Number steps – можно добиться хорошего качества при 60

Stable Diffusion XL (SDXL) — одна из самых сбалансированных архитектур: хорошее качество генерации, минимум встроенной цензуры и относительно умеренные требования к железу. Хотя это уже не самая новая версия, вокруг неё сложилось огромное сообщество — и именно на базе SDXL появилось больше всего специализированных моделей: для аниме, фэнтези, реализма, концепт-арта и т.д.

Писать промпты для SDXL можно так же, как и для SD 1.5: начинайте с общих тегов качества — masterpiece, best quality, 8k, ultra-detailed — они по-прежнему работают.

Один из самых популярных чекпоинтов на SDXL — Pony Diffusion v6 XL. Его особенность — обучение на изображениях, каждое из которых было вручную оценено по шкале от 1 до 10. Эти оценки встроены прямо в промпты, поэтому теперь вы можете «задавать уровень качества» напрямую: просто добавьте score_9, score_8_up или score_7_up — и модель будет стремиться к соответствующей детализации, композиции и чистоте.



Prompt - score_9, score_8_up, score_8, score_7_up, score_7, score_6_up, score_6, score_5_up, Score_5, short male, orange eyes, eyeshadow. makeup, dark orange hair, streaked hair, short cropped hair, fade hairstyle, orange and black horns, huge jagged horns, Tattoos. Grey skin. Gray skin. Dark grey skin, (armor, chestplate, gorget, pauldrons), vox machine style, 1boy, male focus, solo, portrait, cave background, fantasy, (dynamic lighting:1.1), ((masterpiece))

Negative prompt - score_4, pony, 3d, censored, furry, white and black style, long neck, simple background, cleavage

Number steps – на этой модели уже будет оптимальным на 40-45

FLUX

FLUX — современная генеративная модель от Black Forest Labs (2024), созданная бывшими разработчиками Stable Diffusion. Выделяется высоким качеством, точной передачей текста, композицией и пониманием сложных промптов — особенно в реализме и цифровой иллюстрации.

Четыре официальные версии:

FLUX.1 [dev] — самая мощная и качественная; требует лицензию для коммерческого использования (бесплатна для исследований и личного пользования).

FLUX.1 [schnell] — бесплатная, полностью открытая (Apache 2.0), чуть уступает в детализации и сложных сценах, но отлично подходит для локального запуска и дообучения.
FLUX.1 [pro] — облачная, API-only, максимальное качество + расширенные фичи (вроде высокого разрешения и улучшенного inpainting); доступна через платные сервисы (например, Fal, Replicate, Mage.Space).
FLUX.1-Kontext-dev – позволяет редактировать фотографии по комментариям вроде – make hair blue color. Но альтернативные более новые походы справляются с этой задачей лучше

Для художников: FLUX особенно хорош там, где важны пропорции, анатомия, текст и сложные сцены — например, книжные иллюстрации, обложки, концепт-арт. Промпты для неё — проще и короче, чем для SDXL: избыточные теги вроде masterpiece, 8k зачастую не нужны и даже вредят. Для написания промптов для этого модели можно использовать прямую речь и не указывать негативный промпт в отличие от предыдущих моделей. Количество шагов для этой модели 30-35 уже дает отличную детализацию. А так же это первая модель которая может генерировать небольшие надписи.

! ВАЖНО не ставить разрешение стороны выше 1280



Prompt - light-skinned man around 35 years old with black dreadlock hair tied in a bun and a beard like. He is wearing a green and white robe with gold trim, and a blue sash. He is holding a book With text "flux" on it in his left hand and a green orb in his right hand. The style is reminiscent of digital painting with soft, diffused lighting and a focus on character detail.

Далее идут модели которые будет сложно запустить на локальных ПК , но считается что лучшая открытая модель на момент осени 2025 года считается <https://github.com/Tencent-Hunyuan/HunyuanImage-3.0> , для ее запуска потребуется порядка 240-320 гб гпу, что довольно много.

Инструменты доступные на открытых моделях

Коммерческие закрытые модели

К коммерческим (проприетарным) моделям относятся те, что можно использовать лишь на сайте их владельцев либо платно по API без возможности

Есть много платных сервисов с доступом к множеству разных моделей для генерации и редактирования изображений, но я предпочитаю использовать freepik.com, там собраны все

основные лучшие модели и есть безлимитный доступ при достаточно низкой цене и по большинству платных моделей я буду давать инструкции внутри этого сервиса.

Основные подсказки по написанию промтов те же что выше, чем подробнее, тем лучше, стройными грамотными предложениями, желательно до 100 слов, чем длиннее описание не больше деталей будут теряться.

Панель управления выглядит достаточно просто , просто выбираешь модель и пишешь промт свободной речью на английском языке.

The screenshot shows the user interface of an AI application. On the left, there's a sidebar with sections for 'MODEL' (set to 'Google Nano Banana'), 'ОПОРНЫЕ ИЗОБРАЖЕНИЯ' (Style and Character selected), and 'ЗАПРОС' (Request input field). The main area displays a list of models:

- Seedream 4 4K** (Popular) - Described as the only 4K video with reference images and strong aesthetic.
- Seedream** (New) - Described as having unusual creative potential.
- Flux** - Described as a choice of the AI community.
- Mystic** - Described as Freepik AI with 2K resolution.
- Google** - Described as photorealism and matching the request.
- Ideogram 3** - Described as typography and graphic design.
- GPT** - Described as OpenAI technology from ChatGPT.

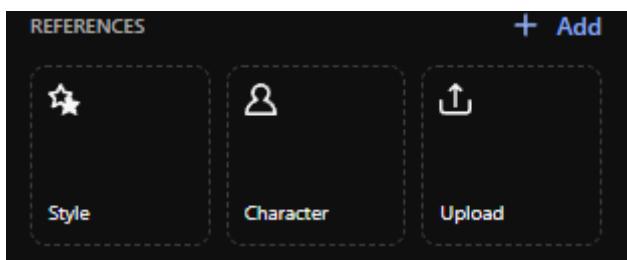
Ниже приведены примеры генерации разных моделей с одним и тем же промтом, на книгах написаны используемые модели.

Из своего опыта могу сказать что:

1. GPT1-HQ рисует максимально детально и соответствует промту, но все картинки он сильно желтит и сразу видно что картинка сгенерирована.
2. Если надо сгенерировать что-то максимально детальное в 4к разрешении в меру креативное то отлично подойдет Seedream 4.0 , особенно хороша для реалистичных изображений
3. Наиболее креативная модель Nano Banana, она свободно чувствует себя в добавлении хороших контекстных деталей, если нет четкого понимания что хочется получить, то лучше всего использовать ее.
4. В генерацию текста на картинке лучше всего все 3 вышеперечисленные модели



Image to Image generation



В сервисе так же можно добавлять своих персонажей и стили, достаточно перетащить референсную картинку в первые 2 поля в зависимости от задачи и модель сама сообразит как это использовать, однако лучше дать небольшое описание главных деталей персонажа для лучшего соответствия. Однако при генерации я предпочитаю третье поле, где можно закинуть любое

свободное изображения (стиль, картинка, деталь, персонаж и т.д.) и напрямую указать взаимодействие с этой картинкой в формате “A digital painting features a fair-skinned woman, around 25 years old like @img1 “ где img1 название приложенного изображения.

MidJorney

Есть официальный гайд от разработчиков <https://docs.midjourney.com/hc/en-us/articles/32040250122381-Image-Prompts> Но из своего опыта скажу что эта модель все еще не научилась понимать прямой речи и требует перечисление словосочетаний через запятую так же как старые SD модели. Генерирует она более художественно, но совершенно не умеет генерировать текст. Ниже приведен пример промта конвертированный из общего описания.



Prompt - digital painting, light-skinned man, 30s, black dreadlocks in a bun, full beard, serene, ornate green and white robe with gold embroidery, deep blue sash — holding an ancient book labeled "MidJourney" in left hand, glowing green orb in right —

Artistic Medium: digital painting

Time Period: Victorian scholar

Emotion: serene, wise

Colors: emerald, ivory, gold, indigo

Environment: old library with shelves of leather-bound books, soft dust motes in light beams