НЕЙРОВИДЕО

ПОЛНОЕ ПРАКТИЧЕСКОЕ РУКОВОДСТВО ПО СОЗДАНИЮ ВИДЕО С ПОМОЩЬЮ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

ИДЕИ
ИНСТРУМЕНТЫ
ПРОЦЕСС
ГОТОВЫЕ ПРИМЕРЫ

СНИМАЙ
БОЛЬШЕ,
ДУМАЙ
ШИРЕ

Милана Панова

НЕЙРОВИДЕО. Полное практическое руководство по созданию видео с помощью искусственного интеллекта

«Автор»
2026

Панова М.

НЕЙРОВИДЕО. Полное практическое руководство по созданию видео с помощью искусственного интеллекта / М. Панова — «Автор», 2026

Три года назад видео из нейросети означало дёрганый ролик, где у людей плыли лица. Сегодня один человек с ноутбуком и подпиской за пару тысяч рублей собирает рекламу, клип или курс — без камеры, студии и монтажёра. Эта книга — рабочее руководство по тому, как делать это в 2026 году.Какие нейросети реально работают сейчас (Sora, Veo, Kling, Runway, Pika и другие), сколько стоят в рублях и как оплатить подписку из России, если сервис не принимает российские карты. По какому принципу рождается видео внутри модели и куда всё продвинулось. Небанальный промтинг на языке кино, подготовка референсов, из которых собирается качественная картинка, длительность и число роликов за раз, расход кредитов. Отдельно — нейросети для озвучки и полный пайплайн от идеи до готового ролика.Все цены и способы оплаты сверены на 2026 год. Книга для блогеров, маркетологов, SMM-специалистов, предпринимателей и всех, кто хочет делать видео с помощью ИИ — от первого ролика до серийного контента.

Содержание

Милана Панова
НЕЙРОВИДЕО. Полное практическое руководство по созданию видео с помощью искусственного интеллекта

НЕЙРОВИДЕО
2026
Полное практическое руководство по созданию видео
с помощью искусственного интеллекта
Актуальные нейросети · цены в рублях · оплата из России
промтинг · референсы · озвучка · полный пайплайн
Издание обновлено: сентябрь 2026

От автора и о том, как читать эту книгу

Ещё три-четыре года назад «видео из нейросети» означало пятисекундный ролик, где лица плыли, у людей вырастал шестой палец, а вода вела себя как желе. Сегодня один человек с ноутбуком и подпиской за пару тысяч рублей в месяц собирает рекламный ролик, короткометражку, обучающий курс или серию контента для соцсетей — без камеры, студии, актёров и монтажёра. Технология прошла путь, на который у классического кино ушли десятилетия, за считаные годы.

Эта книга — не сборник восторгов и не рекламная брошюра одного сервиса. Это рабочее руководство. Она отвечает на конкретные вопросы: какие нейросети реально работают в 2026 году, сколько они стоят в рублях, как их оплатить из России, по какому принципу вообще рождается видео внутри модели, как писать промты, которые дают результат, где брать и как готовить референсы, сколько роликов и какой длины можно получить за один раз, во что это обходится в кредитах и чем озвучивать готовое видео.

Важно про актуальность. Рынок ИИ-видео меняется буквально каждый месяц: выходят новые версии моделей, меняются цены, тарифные сетки и способы оплаты. Все цифры в книге сверены по открытым источникам на середину–конец 2026 года и указаны как ориентир. Перед оплатой всегда проверяйте актуальную цену на официальной странице сервиса или у выбранного посредника — это единственный источник истины на день покупки.

О ценах в рублях. Почти все зарубежные сервисы берут оплату в долларах. Рублёвые суммы в книге рассчитаны по ориентировочному курсу около 95–100 ₽ за $1 и уже учитывают, что при оплате через посредников и виртуальные карты сверху добавляется комиссия примерно 5–15%. Реальная сумма всегда зависит от курса на день оплаты и выбранного способа, поэтому воспринимайте рублёвые цифры как «плюс-минус», а не как прайс.

Книга построена так, чтобы её можно было читать подряд — от теории к практике — или открывать нужную главу как справочник. Новичку я советую пройти всё по порядку. Тому, кто уже что-то генерировал, — сразу листать к главам про промтинг, референсы и пайплайн.

Содержание

Внутри каждой главы — тематические разделы; удобная навигация по ним доступна в оглавлении вашего ридера.

Глава 1. Как нейросеть создаёт видео: от глитчей до кинематографа

1.1. Как было раньше: эпоха склеенных кадров

Первые попытки заставить нейросеть «нарисовать» видео опирались на архитектуру GAN — генеративно-состязательные сети. Идея красивая: одна сеть (генератор) создаёт кадр, вторая (дискриминатор) пытается отличить подделку от настоящего изображения, и в этой борьбе генератор постепенно учится делать правдоподобно. Для отдельных картинок это работало неплохо, но видео — это не картинка, а десятки картинок в секунду, которые обязаны быть согласованы между собой.

Именно на согласованности всё и ломалось. Модель генерировала кадры почти независимо, и человеческий глаз мгновенно ловил несоответствия: лицо «дышало» и меняло черты, фон подрагивал, предметы возникали и исчезали, руки превращались в кашу из пальцев. Ролики были короткими — две-четыре секунды, — потому что чем дальше от первого кадра, тем сильнее модель «забывала», что было в начале.

Переходным этапом стали диффузионные модели для видео вроде **Stable Video Diffusion**. Логика была такая: берётся один ключевой кадр, и от него достраивается короткая последовательность на 2–4 секунды с последующей стабилизацией и апскейлом. Уже лучше, но всё ещё «оживление картинки», а не полноценная сцена с логикой и движением камеры.

1.2. Диффузия: как из шума рождается изображение

Современное ИИ-видео почти целиком построено на диффузии. Это стоит понять один раз — тогда станет ясно, почему промты работают именно так и откуда берутся кредиты и время генерации.

Прямая диффузия (зашумление). На этапе обучения модели берут настоящее изображение или видео и шаг за шагом подмешивают в него случайный шум, пока от исходника не останется ничего — просто «телевизионная рябь». Модель наблюдает этот процесс на миллионах примеров и запоминает, как выглядит разрушение картинки на каждом шаге.

Обратная диффузия (расшумление). А теперь самое интересное: модель учится идти в обратную сторону — из чистого шума постепенно, за десятки шагов, «проявлять» осмысленное изображение, как фотография в проявителе. Когда вы нажимаете «Сгенерировать», сеть стартует именно от случайного шума и шаг за шагом лепит из него кадры, ориентируясь на ваш текст.

Отсюда сразу два практических следствия. Во-первых, чем больше шагов расшумления, тем чище и детальнее результат — но тем дороже и дольше генерация (быстрые режимы вроде Turbo/Lightning жертвуют деталями ради скорости и предпросмотра). Во-вторых, результат управляется на каждом шаге, поэтому текст промта, стартовая картинка и дополнительные сигналы реально «рулят» тем, что проявится из шума.

1.3. Латентное пространство: почему это вообще возможно

Расшумлять честное видео в 4K по пикселям было бы слишком дорого даже для дата-центров. Поэтому модели работают не с пикселями, а в **латентном пространстве** — сжатом математическом представлении, где картинка хранится как компактный код, а не как миллионы точек. Слово «латентный» означает «скрытый»: модель сначала планирует и собирает сцену «в уме», в этом сжатом виде, и только в самом конце декодер разворачивает код обратно в настоящие пиксели.

Именно это дало рывок: работать со сжатым представлением на порядок дешевле, поэтому стали возможны и высокое разрешение, и большая длина, и приемлемая скорость.

1.4. Diffusion Transformer: архитектура, которая всё изменила

Долгое время «мозгом» диффузионных моделей была архитектура U-Net. Прорыв случился, когда её заменили на **трансформер** — ту самую архитектуру, что стоит за большими языковыми моделями. Получился **латентный диффузионный трансформер (Diffusion Transformer, DiT)**. На нём построена Sora, и по этому же пути пошли остальные флагманы.

Ключевая идея DiT: видео разбивается на «патчи» — маленькие кусочки пространства и времени, — и трансформер обрабатывает их так же, как языковая модель обрабатывает слова в предложении. Он видит связи между кусочками не только внутри одного кадра, но и между кадрами во времени. Отсюда — понимание физики, устойчивость сцены и связность движения, которых так не хватало старым моделям.

У трансформеров есть ещё одно волшебное свойство — **масштабируемость**: чем больше модель и чем больше данных, тем предсказуемо лучше результат. Это превратило улучшение видео-ИИ из череды случайных удач в понятную инженерную задачу «добавь вычислений и данных» — и объясняет, почему прогресс пошёл лавиной.

1.5. Полный путь от вашего промта до кадра

Соберём всё вместе. Когда вы отправляете запрос, внутри происходит примерно следующая цепочка:

Кодирование текста. Ваш промт превращается в набор чисел (эмбеддинги), понятных модели, — это «техзадание» на языке сети.

Старт от шума. В латентном пространстве создаётся случайный шум нужного размера — заготовка под будущее видео.

Диффузия под управлением текста. За десятки шагов сеть расшумляет заготовку, на каждом шаге сверяясь с вашим промтом (а если вы дали стартовую картинку — то и с ней).

Декодирование. Готовый латентный код разворачивается декодером в настоящие пиксели — последовательность кадров.

Постобработка. Кадры собираются в ролик, при необходимости апскейлятся, стабилизируются, к ним добавляется звук.

1.6. Куда продвинулись к 2026 году и почему это важно

За последний виток развития видео-ИИ научился нескольким вещам, которые ещё недавно казались фантастикой:

- **Разрешение и плавность.**
  Кинематографичные сцены в 4K с частотой до 60 кадров в секунду вместо дёрганых 480p.
- **Длина.**
  От 2–4 секунд у ранних моделей до 15–25 секунд одним куском у флагманов, а с монтажной склейкой — до нескольких минут связного повествования.
- **Нативный звук.**
  Veo, Sora и Kling умеют генерировать видео сразу со звуком: диалоги, эффекты, фоновый эмбиент и синхронизация губ с речью — без отдельной озвучки.
- **Консистентность персонажа.**
  Технологии вроде Runway с референсами и «Character»-систем позволяют сохранить одно и то же лицо и одного и того же героя между разными сценами — основа сериального контента.
- **Физика.**
  Вода течёт как вода, ткань развевается на ветру, тени и отражения ведут себя правдоподобно.
- **Управление камерой.**
  Motion Control и режиссёрские приёмы: наезд, панорама, орбита, кран, слежение — прямо из промта или по заданной траектории.
- **Мультикадр.**
  Multi-Shot-режимы собирают внутри одной генерации несколько планов одной сцены.

Почему рывок случился именно сейчас — это сумма факторов: переход на масштабируемые трансформеры, работа в дешёвом латентном пространстве, огромные датасеты «видео + описание», инженерные оптимизации скорости и, конечно, жёсткая конкуренция OpenAI, Google, Kuaishou, Runway и китайских лабораторий, которая гонит всех вперёд.

Глава 2. Карта актуальных нейросетей для видео (2026)

Главный вывод рынка 2026 года: универсального победителя нет. Вопрос звучит не «какая нейросеть лучшая», а «какая лучше под мою конкретную задачу». Одни сильны в реалистичных пейзажах, другие — в людях и движении, третьи — в аватарах, четвёртые — в стилизации. Ниже — карта основных игроков, сгруппированных по назначению.

Конец ознакомительного фрагмента.