

Генерация видео по текстовому описанию с использованием нейросети Kling Elements

Екимова Яна Сергеевна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Целью работы является создание видеоролика, точно соответствующего текстовому описанию (промту), который демонстрирует динамичное взаимодействие между главным героем и фантастическим существом в естественной окружающей среде. Генерация осуществляется с помощью нейросетевой модели Kling Elements с ключевыми параметрами: разрешение 4К, высокая частота кадров и кинематографическое качество освещения и анимации для достижения максимальной реалистичности и визуального воздействия. Для реализации поставленной задачи была выбрана нейросетевая модель Kling Elements, способная точно интерпретировать сложные промты, генерировать плавную и физически достоверную анимацию движений, а также обеспечивать детализированную проработку текстур и окружающей среды. Этот подход позволяет не только создавать качественный контент для анимации и кинематографии, но и служит мощным инструментом для превизуализации (pre-visualization) в профессиональных кинопроизводственных пайплайнах.

Ключевые слова: Kling Elements, нейросеть, генерация видео.

Video generation based on a text description using the Kling Elements neural network

Ekimova Yana Sergeevna

Sholom Aleichem Priamurskiy State University

Student

Abstract

The aim of the work is to create a video clip that accurately matches the text description (prompt), which demonstrates the dynamic interaction between the main character and a fantastic creature in a natural environment. The generation is carried out using the Kling Elements neural network model with key parameters: 4K resolution, high frame rate and cinematic quality of lighting and animation to achieve maximum realism and visual impact. To achieve this goal, the Kling Elements neural network model was chosen, capable of accurately interpreting complex artifacts, generating smooth and physically accurate motion animations, as well as providing detailed study of textures and the environment. This approach allows not only to create high-quality content for animation and cinematography,

but also serves as a powerful tool for pre-visualization in professional film production pipelines.

Keywords: Kling Elements, neural network, video generation.

1 Введение

1.1 Актуальность

Современный этап развития искусственного интеллекта характеризуется стремительным ростом технологий генеративного моделирования, особенно в области визуального контента. В условиях высокого спроса на персонализированный и динамичный видеоматериал в медиа, рекламе, образовании и развлечениях, системы генерации видео по описанию становятся ключевым инструментом цифровой трансформации. Среди них особое место занимает Kling Elements — мощная нейросетевая модель, способная создавать высококачественные, реалистичные видео по текстовым запросам с учётом временной согласованности, плавности движений и детализации сцены.

1.2 Обзор исследований

Д.В. Фисько посвятил статью исследованию возможностей условных генеративных нейросетевых моделей для создания изображений высокого разрешения и их практического применения в разработке инновационных продуктовых решений. В работе систематизирован исторический путь развития методов условной генерации, проанализированы современные архитектуры нейросетевых моделей и методы их обучения [1]. Генеративно-состязательные сети для генерации изображений по тексту использовали А. О. Левин, Ю. С. Белов [2]. И. С. Шустов, Ю. С. Белов подготовили и обработали данные в задачах генерации фотореалистичных изображений [3]. В статье В.В. Кабанова рассматривается современное состояние и возможности искусственного интеллекта в области манипуляций с изображениями человеческих лиц [4]. И.Б. Навойчик, рассмотрел применение искусственного интеллекта (ИИ) для генерации медиаконтента [5].

1.3 Цель исследования

Целью исследования является генерация видео, визуализирующее сцену тёплого игрового взаимодействия между ребёнком и фантастическим существом (пикачу) в естественной природной среде, с использованием текстового промта и нейросетевой модели Kling Elements, обеспечив высокое качество изображения (4K), реалистичную детализацию окружающей среды и гармоничное сочетание реального и фантастического элементов.

2 Материалы и методы

В данном исследовании используется нейросетевая модель Kling Elements, реализованная на платформе Gen-API, предназначенная для генерации видео по текстовому описанию.

3 Результаты

Переходим на сайт <https://gen-api.ru/model/pixverse-v4>, в разделе «Нейросети» выбираем модель Kling Elements.

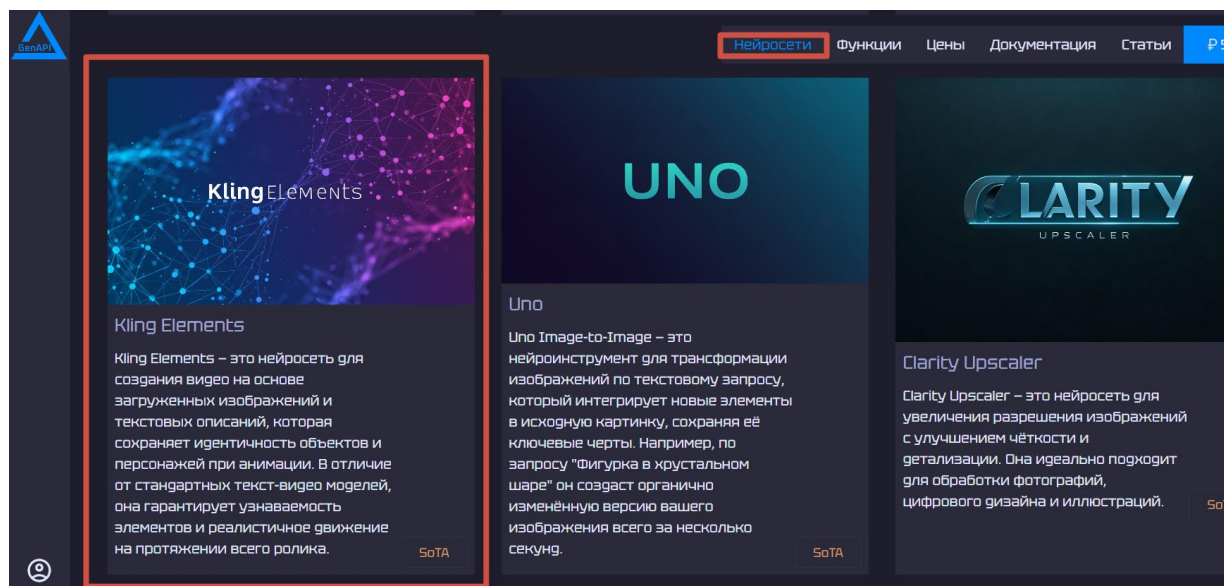


Рисунок 1. Выбор модели

После того, как выбрали нужную модель переходим в поле ввода промта (prompt) и пишем нужный текст: «Мальчик играет с Пикачу в красивом парке среди высоких елей. Вокруг - зелёная трава, солнечные лучи пробиваются сквозь хвою, лёгкий ветерок. Атмосфера волшебная и тёплая. Высокое качество, 4К.»

На следующем этапе осуществляется загрузка исходных изображений, необходимых для последующей генерации контента с использованием нейросетевой модели.

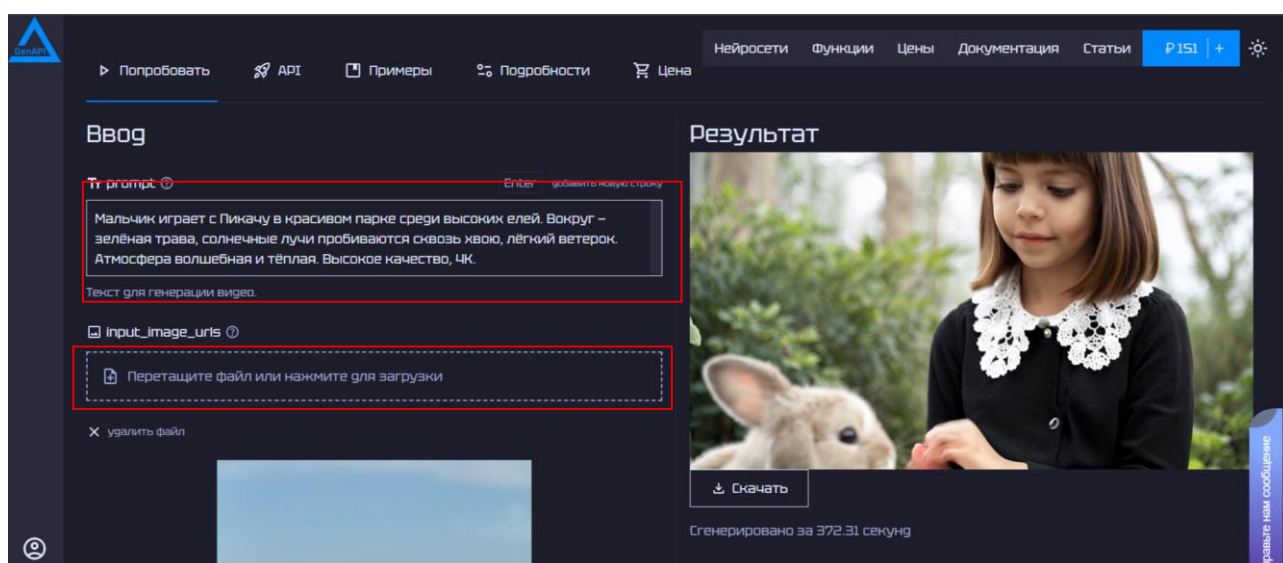


Рисунок 2. Ввод данных

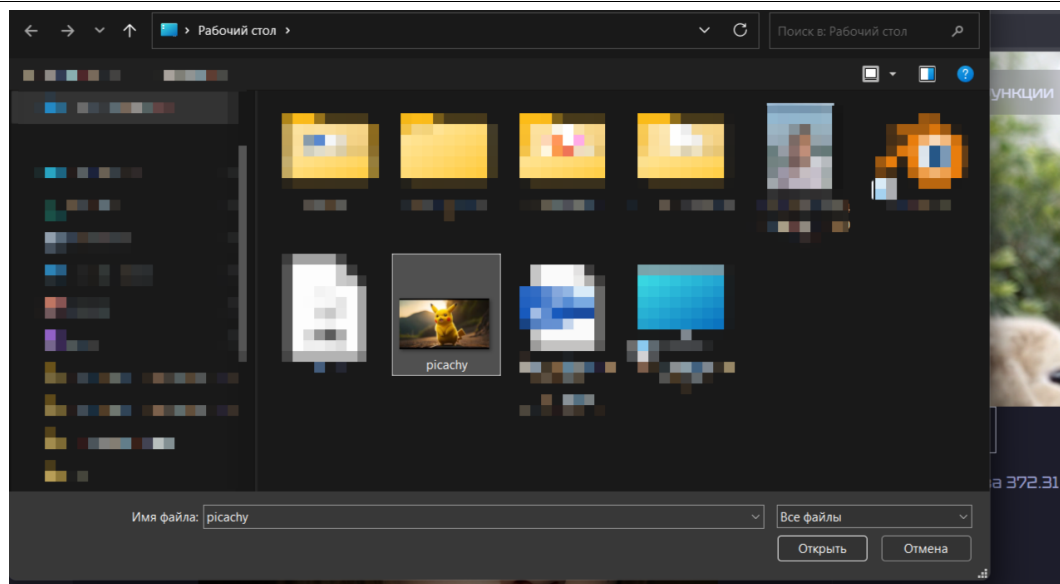


Рисунок 3. Загрузка фотографии

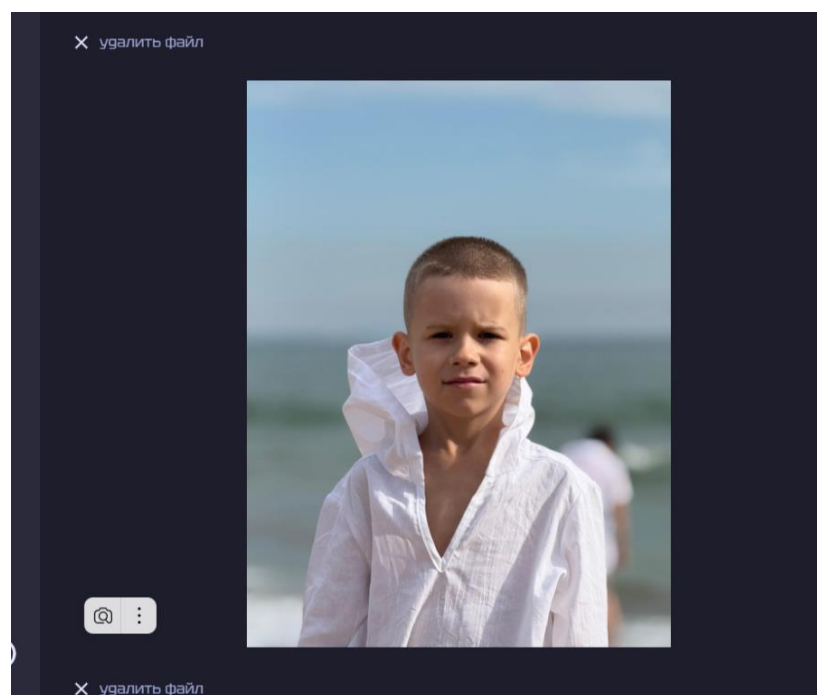


Рисунок 4. Загруженная фотография

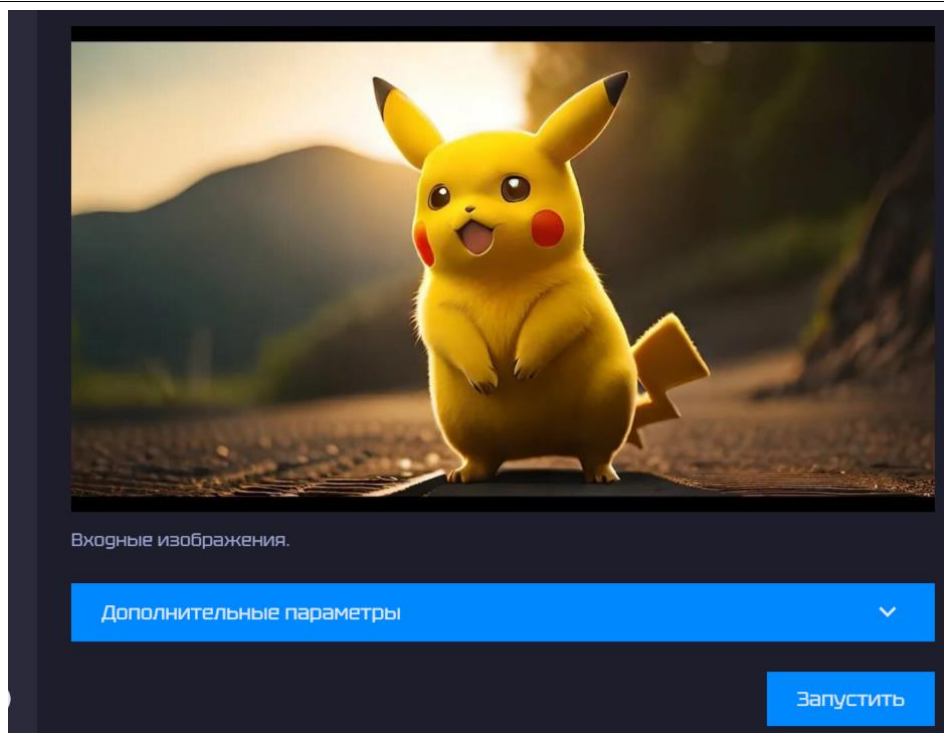


Рисунок 5. Загруженная фотография

После нажатия кнопки «Запустить» инициируется процесс генерации видео, который завершается автоматически после выполнения всех вычислительных этапов.

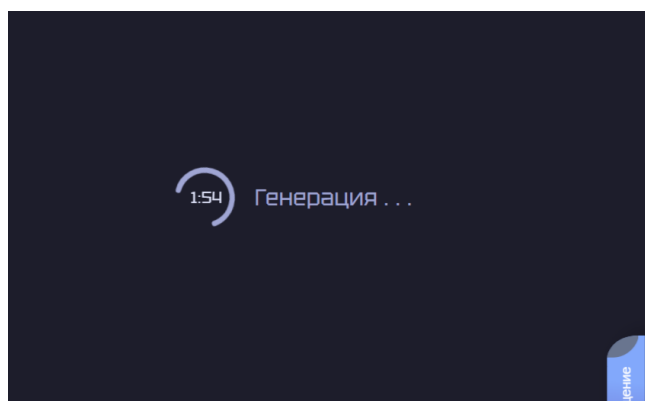


Рисунок 6. Процесс генерации

После того, как закончится генерация, система предоставит готовый видеоролик, который можно загрузить на устройство с помощью кнопки скачивания. Готовое видео можно скачать по ссылке: <https://disk.yandex.ru/i/QINu0oZG-NjSDw>.

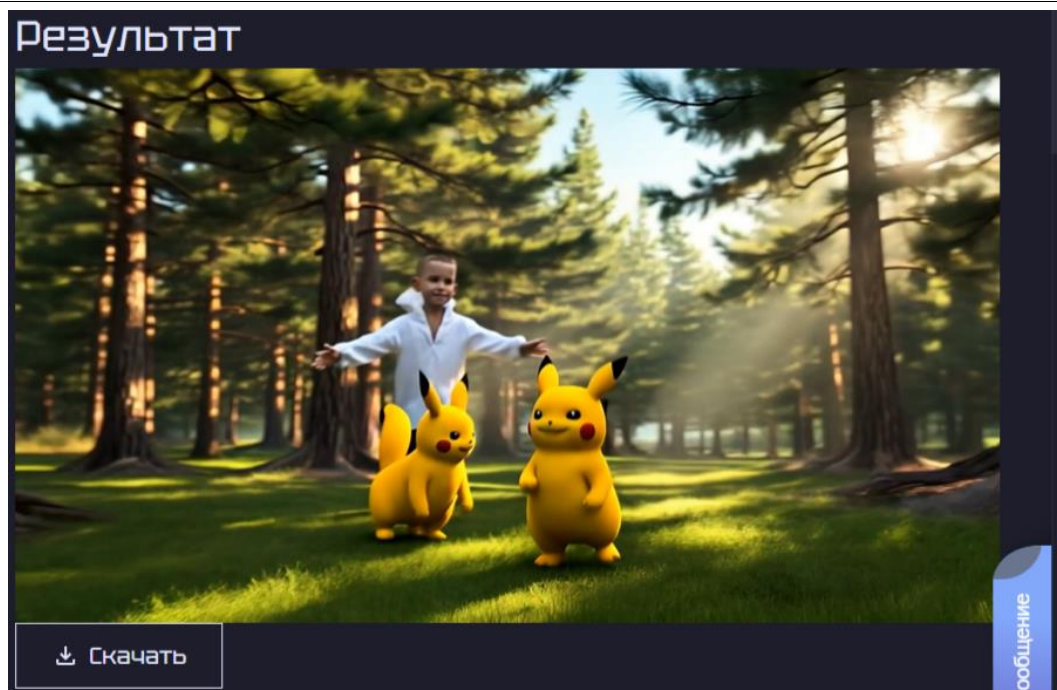


Рисунок 7. Готовое видео

4 Выводы

В данной работе подробно рассмотрен процесс генерации видео с использованием современной нейросетевой модели Kling Elements, демонстрирующей высокий уровень синтеза визуального контента на основе текстовых промтов. Были описаны ключевые этапы создания видео: формулирование детализированного запроса, загрузка вспомогательных материалов (при необходимости), запуск генерации и последующее скачивание результата. Показаны преимущества платформы, включая высокое качество выходного видео (до 4K), поддержку реалистичной анимации, плавных переходов и временной согласованности кадров.

Библиографический список

1. Фисько Д. В. Обзор методов условной генерации изображений нейросетевыми моделями // Актуальные вопросы современной науки и технологий. 2021. С. 57-62.
2. Левин А. О., Белов Ю. С. Использование генеративно-сопоставительных сетей для генерации изображений по тексту // Научное обозрение - технические науки. 2023. №2. С. 10.
3. Шустов И. С., Белов Ю. С. Подготовка и обработка данных в задачах генерации фотореалистичных изображений // Международный студенческий научный вестник. 2018. №. 5. С. 139-139.
4. Кабанова В. В. Генерация и трансформация лица на изображении с использованием искусственного интеллекта // Регион искусственного интеллекта: Материалы Всероссийской студенческой научной конференции, Череповец, 06 ноября 2022 года. Череповец: Череповецкий государственный университет, 2022. С. 33-37.

5. Навойчик И. Б. Применение искусственного интеллекта при генерации медиаконтента // Электронные системы и технологии : сборник материалов 61-й научной конференции аспирантов, магистрантов и студентов БГУИР, Минск, 21–25 апреля 2025 г.. Белорусский государственный университет информатики и радиоэлектроники. Минск, 2025. С. 853–856.