

# Jukebox: A Generative Model for Music

Prafulla Dhariwal Heewoo Jun Christine Payne Jong Wook Kim Alec Radford Ilya Sutskever

Автор обзора-рецензии:

Добряев Иван

## 1. Суть работы и основной вклад:

Работа посвящена модели Jukebox, которая моделирует музыку в виде необработанного аудио сигнала. Для этого команда Open AI используем многоуровневую VQ-VAE-архитектуру, чтобы сжать аудио до дискретных кодов, а также моделирует их, используя авторегрессионные трансформеры. Авторы показали, что их комбинированная модель может генерировать качественную и разнообразную музыку с коэффициентом связности, достигающим нескольких минут (раньше такой длины не получалось достичь).

## 2. Авторы:

*Prafulla Dhariwal*

Researcher, OpenAI  
h-index 15  
NLP, CV, Loss function etc.

*Jong Wook Kim*

Researcher, OpenAI  
h-index 11  
Music Information Retrieval

*Alec Radford*

Researcher, OpenAI  
h-index 31  
NLP

*Ilya Sutskever*

Co-Founder and Chief Scientist of OpenAI  
h-index 75

## 3. Основные статьи:

### 1. Neural Discrete Representation Learning

## 4. Цитирование и продолжение:

Msanii: High Fidelity Music Synthesis on a Shoestring Budget 2023 (WIP) diffusion-based model.

AudioLM: a Language Modeling Approach to Audio Generation Google

У данной статьи нет прямого продолжения. А также цитирование происходит как указание существующих других подходов, а не заимствование подхода.

## 5. Сильные стороны:

Новый на то время подход, что дает возможность сгенерировать достаточно высококачественную и разнообразную музыку с корреляцией, способной продержаться в течение нескольких минут. Также можно управлять стилем музыки и вокала с помощью текстов и тд.

## 6. Слабые стороны:

Результаты экспериментов демонстрируют достаточно высокое качество в генерации песен (со слов авторов), но в тоже время нет достаточных данных для понимания и измерения того, насколько генерируемые результаты соответствуют музыке, созданной человеком и отсутствует сравнение с предыдущими подходами на более коротких интервалах времени.

## 7. Улучшения:

Для улучшения текущей работы можно было бы исследовать влияние внешних, не разговорных, форм на музыкальные результаты, такие как паузы. Также стоит проанализировать влияние контроля по лирике и художественной области музыки на адекватность генерируемой музыки.