

Riffusion is an innovative neural network developed by Seth Forsgren and Hayk Martiros, which produces music by utilizing visual representations of sound, known as spectrograms, rather than traditional audio files. This technology is an advanced adaptation of the pre-existing open-source platform, Stable Diffusion, which is primarily used for creating images from textual prompts. Riffusion operates by transforming text prompts into image files, which can subsequently undergo an inverse Fourier transform to be converted into short audio clips. Additionally, the model has the capability to blend different audio files seamlessly by leveraging the latent space between outputs, a feature derived from the img2img function of the Stable Diffusion model.

The music generated by Riffusion has been characterized as surreal and unique, though it is not expected to supplant human-composed music. Released on December 15, 2022, Riffusion's code is openly accessible on GitHub, adding to the growing list of applications based on the Stable Diffusion model.

As part of the emerging field of AI-driven text-to-music generators, Riffusion shares the space with other notable projects. In December 2022, Mubert employed Stable Diffusion to transform descriptive text into musical loops. Moreover, in January 2023, Google introduced MusicLM, their contribution to the text-to-music generator landscape, through a published research paper.