



## The Mixology Dataset

Michael Clemens <sup>1</sup>

<sup>1</sup> University of Utah

Published: 04 November 2023

### License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

### Abstract

Generative artificial intelligence (AI) is making strides in AI music, but its effectiveness largely depends on the diversity of audio samples from various genres. The transparency of the datasets used by many successful models has sparked concerns among artists about the legality of the music data used for training. As generative AI models require more data for improved performance, the challenge lies in ethically sourcing this data while adhering to the music industry's licensing rules.

While there is a significant need for large, open-source audio datasets for training, the scarcity of data for mixing and mastering tasks in music production is even more pronounced. Parameter values are often confined within the digital audio workstation and not made publicly available. We introduce **The Mixology Dataset** to address these gaps, an extension of Brecht De Man et al.'s *The Mixing Evaluation* dataset ([De Man & Reiss, 2017](#)). Our dataset offers individual track and plugin settings for 114 mixes spanning 34 songs. It includes details such as *gain*, *pan*, *equalization*, *compression*, and *reverb* settings, along with the raw audio for each track, all structured within a JSON format.

The licensing terms of each song are maintained, and parameter values were collected only from non-copyrighted songs. We recommend using the AI2 ImpACT-LR License ([AI2 ImpACT Licenses](#)) for all artifacts generated using this dataset. Additionally, we provide a tool that can systematically extract parameter values from images of plugin screenshots. This dataset affords the ability to separate parameter recommendations by individual tracks such as drums, bass, vocals, etc. Although this dataset is not intended solely for audio source separation, it can be used to assist in source separation for parameter recommendation by track type. We hope this contribution will aid in developing co-creative agents for assisted mixing and mastering.

AI2 ImpACT licenses. <https://allenai.org/impact-license>.

De Man, B., & Reiss, J. D. (2017). The mix evaluation dataset. *20th Int. Conf. On Digital Audio Effects (DAFx-17)*.