



# Ethics & aesthetics towards building an artistic discourse around music AI.

—

@jordiponsdotme - [www.jordipons.me](http://www.jordipons.me)

May 2021



## DISCOURSE.

~~Music AI: garbage in, garbage out.~~

Discourse is the result of study, reflection and intention.

It can be challenging for a non-technical person.

Studying current music AI we discover its **AESTHETICS** and **ETHICS**.



Local motives  
[Wavenet samples](#)



Artifacts  
[Jukebox samples](#)

---

## AESTHETICS.

~~Music AI: solved problem.~~

Two relevant challenges: capture long-term dependencies, and high-fidelity audio synthesis.

Aesthetical facets of music AI in 2010s: **"LOCAL MOTIVES"** and **"ARTIFACTS"**.

Aesthetics of **"PERFECTION"**: in plausible compositions given a specific style or music tradition.

- Example of plausible AI composition: [The AI Music Generation Challenge 2020](#)





Local motives  
[Wavenet samples](#)



Artifacts  
[Jukebox samples](#)

---

## AESTHETICS.

~~Music AI: solved problem.~~

Two relevant challenges: capture long-term dependencies, and high-fidelity audio synthesis.

Aesthetical facets of music AI in 2010s: “**LOCAL MOTIVES**” and “**ARTIFACTS**”.

Aesthetics of “**PERFECTION**”: in plausible compositions given a specific style or music tradition.

- Example of plausible AI composition: [The AI Music Generation Challenge 2020](#)



## ETHICS.

*Is ethical to imitate an artist? Who owns the copyright of this new piece?*

*Is ethical to generate infinite tunes on a style?*

*Have you thought about cultural ownership/appropriation?*

*What's the role of the artist if songs are generated by an AI?*

Current music AI is a data-driven process triggered by **INTENTIONED** humans.



# Ethics & aesthetics towards building an artistic discourse around music AI.

—

@jordiponsdotme - [www.jordipons.me](http://www.jordipons.me)

May 2021

—  
**Discourse.**



# Holly Herndon: on embodiment, data, ethics and aesthetics

On **embodiment**: "She (Spawn) is our AI baby"

On **data and ethics**:

"At the end of the day, AI is just us. That's the dirty secret of AI: you have this bast  
amount of human intelligence that goes into the training datasets"

They curate they own datasets with artistic intention: Holly's voice, recordings of a choir, or even their public in live shows.

"We wanted to give attribution to those who went into training spawn"

On **aesthetics**:

"We were surprised of how lo-fidelity it sounded"

"For now, it has to be textural sound"

"AI contributes to the color of the record"

"AI as a performer, a member of the ensemble"

"Accommodate the aesthetics of AI with the rest of the production"

Extracted from those videos:

[youtube.com/watch?v=k67IKMXzD7A](https://youtube.com/watch?v=k67IKMXzD7A) [youtube.com/watch?v=v\\_4UqpUmMkg](https://youtube.com/watch?v=v_4UqpUmMkg)

# YACHT: on using music AI as a methodology

On **reflecting current society**: "We live in a world where we're completely surrounded by semi-autonomous intelligent entities,"  
"And so as artists, we're interested in making art about the real world that we live in because we felt like there was this really interesting opportunity to understand more about ourselves."

On **dataset creation**: "YACHT worked with LA artist [Ross Goodwin](#) to create their lyric generating algorithm.  
They built a collection of more than 2 million words. The words were drawn from their back catalog,  
their favorite bands, and music they heard growing up."

On **AI as a methodology**: "We decided we were going to be very strict about how we used this material.  
And we created a set of rules. We can't add anything. We can't improvise anything. We can't harmonize,"  
"We decided it would be just a subtractive process. So we could remove things, like we could take out a word,  
but we couldn't add a word for the lyrics. Same with the drum patterns and the melodies."

Extracted from this [interview](#)

# Sony: on novel tools for music creation

“Through deep learning-based approaches, we develop **tools that enhance a composer’s creativity and augment his capabilities**”

“We are often experiencing a **gap** that **exists between scientific research and the music industry** on many levels, such as timeliness or profitability.”

ICASSP2020 Sony Workshop: [2020.ieeeicassp.org/program/workshops/sony-workshop/index.html](https://2020.ieeeicassp.org/program/workshops/sony-workshop/index.html)

“Sony have tools that help **generate rhythms & melodies**, and others that focus purely on **sound design**.”

“I think of the people who helped **developing the MOOG synthesizers or the first drum machines**, musicians were needed to develop them.”

“**It’s not a robot that makes music by itself, it’s really far from that**. In the future, it’s going to provide ways to produce new tools and give novel ideas.”

Uele Lamore's collaboration with Sony: [youtube.com/watch?v=H12OA2rjWeg](https://youtube.com/watch?v=H12OA2rjWeg)

# Bob Sturm and collaborators: on ethics and aesthetics

"My goal is **not to build machines that would compose billions of tunes that the world doesn't need**, and it's not to put people out of business – but it is to **engage with this technology in a meaningful** way and to seek ways of avoiding unintended harm from AI whether it's applied on music or other areas."

The AI Music Generation Challenge 2020 [youtube.com/watch?v=KSoSyoEx6hc](https://youtube.com/watch?v=KSoSyoEx6hc)

**Work closely with the music communities around those traditions** to explore: "How music AI research can benefit from considering traditional music? How traditional music can benefit from music AI research?".

The AI Music Generation Challenge 2021 [github.com/boblsturm/aimusicgenerationchallenge2021](https://github.com/boblsturm/aimusicgenerationchallenge2021)

Look for plausible compositions in a specific style / music tradition, note its evaluation criteria (**aesthetics**):

"Four tunes are selected at random and sent to all judges. If plagiarism detected, reject. If meter is not characteristic of a slängpolska, reject. If rhythm is not characteristic of a slängpolska, reject. Judges also rate the following qualities: danceability, stylistic coherence, formal coherence, and playability"

The AI Music Generation Challenge 2021 [github.com/boblsturm/aimusicgenerationchallenge2021](https://github.com/boblsturm/aimusicgenerationchallenge2021)

## "sampling" style

### **raw samples**

artifacts can appear

### **curated samples**

seeking for realistic renditions

## **interactive sampling**

interact with AI, it can be useful for live shows - but also to tailor the model towards an specific artistic "intention"

# Keyon Christ: on “sampling” AI for hip-hop

“ While rap in essence is far from conservative, when it comes to new technologies like NFTs and AI, the genre has been lagging behind. Repurposing records has been the basis of hip-hop for decades, **but this rich sampling tradition never migrated beyond its traditional confines** - *What do you mean you created hip-hop from the discography of an Artificial Intelligence death metal band?* ”

“**What was the AI-collaborative process like?** Two things. The process was not more difficult than sampling a human song. The process was actually more inspirational being that we had 12 hours of variations on the human element of soul to pull from, there was a continuum of ghostlike waveforms that we could peer into, extract, cut up and make into a coherent ghetto futuristic entanglement.”

On discourse: “The two **(music and AI research) need to converge**. I’m here (and DADABOTS is here) to see to it that the two create a beautiful and ignorant fusion.”

All from this interview: [highsnobiety.com/p/the-worlds-first-nft-rap-track-is-here-weve-got-questions/](https://highsnobiety.com/p/the-worlds-first-nft-rap-track-is-here-weve-got-questions/)

# On extending the “sampling” culture in hip-hop

“Sampling is the reuse of a portion (or sample) of a sound recording in another recording.”

from Wikipedia

Sample pack on artificial intelligence by Andrew Huang:

<http://andrewhuang.com/store/artificial-intelligence-sample-pack>

Think of it as **sampling samples of a distribution** to extend the sampling culture in hip-hop

Even though the strong culture of sampling on hip hop, **Jay-Z** felt uncomfortable with the use of his voice for a generative model:

[theguardian.com/music/2020/apr/29/jay-z-files-takes-action-against-deepfakes-of-him-rapping-hamlet-and-billy-joeel...](https://theguardian.com/music/2020/apr/29/jay-z-files-takes-action-against-deepfakes-of-him-rapping-hamlet-and-billy-joeel...)

# AI to generate what?

**Scores.      MIDI.      Lyrics.      Waveforms.**

See the Eurovision AI Song Contest, as a variate example of  
how people take different approaches to music AI:

[youtube.com/watch?v=-ylu5VLZj5g](https://youtube.com/watch?v=-ylu5VLZj5g)

See this AI music gallery by NVIDIA:

[nvidia.com/en-us/deep-learning-ai/ai-art-gallery/](https://nvidia.com/en-us/deep-learning-ai/ai-art-gallery/)



---

# Aesthetics.

Audio, visual or audio-visual aesthetics?  
Mostly audio.

# Dadabots

“Music genres like **metal and punk seem to work better, perhaps because the strange artifacts of neural synthesis** (noise, chaos, grotesque mutations of voice) are aesthetically pleasing in these styles.”

Generating Albums with SampleRNN to Imitate Metal, Rock, and Punk Bands ([arxiv.org/pdf/1811.06633.pdf](https://arxiv.org/pdf/1811.06633.pdf))

“We were delighted by the **aesthetic merit of the imperfections**. Pioneering artists can exploit these effects, just **as they exploit vintage sound production** (tube warmth, tape-hiss, vinyl distortion, etc). ”

Generating Albums with SampleRNN to Imitate Metal, Rock, and Punk Bands ([arxiv.org/pdf/1811.06633.pdf](https://arxiv.org/pdf/1811.06633.pdf))

“While we set out to achieve a realistic recreation of the original data, we were delighted by the aesthetic merit of its imperfections. Solo vocalists become a lush choir of ghostly voices, rock bands become crunchy cubist-jazz, and cross-breeds of multiple recordings become a surrealist chimera of sound. ”

Generating Black Metal and Math Rock: Beyond Bach, Beethoven, and Beatles ([arxiv.org/pdf/1811.06639.pdf](https://arxiv.org/pdf/1811.06639.pdf))

AI Music Creativity 2020 Spotlight: [youtube.com/watch?v=CiTz5E6rGEg](https://youtube.com/watch?v=CiTz5E6rGEg)

# Local motives

on AI for symbolic music



Local motives  
[LSTM blues samples](#)

from LSTMs to Transformer!



More global motives  
[Music Transformer samples](#)

# 2010 aesthetics

on AI for waveform generation



Local motives  
[Wavenet samples](#)

from Wavenet to Jukebox!



Artifacts  
[Jukebox samples](#)

# Artifacts in neural audio synthesis

Despite the novelty of being able to generate generally high fidelity and coherent songs, sample quality was still limited by a number of factors. First, the use of 22 kHz sampling rate along with small upsamplers introduced noise both in the upsampling and decoding steps, which we hear as grainy texture. We improved fidelity by using 44 kHz VQ-VAE and 1B parameter upsamplers in all subsequent experiments at the expense of longer rendering time.

Jukebox: A Generative Model for Music:  
[arxiv.org/pdf/2005.00341.pdf](https://arxiv.org/pdf/2005.00341.pdf)

## UPSAMPLING ARTIFACTS IN NEURAL AUDIO SYNTHESIS

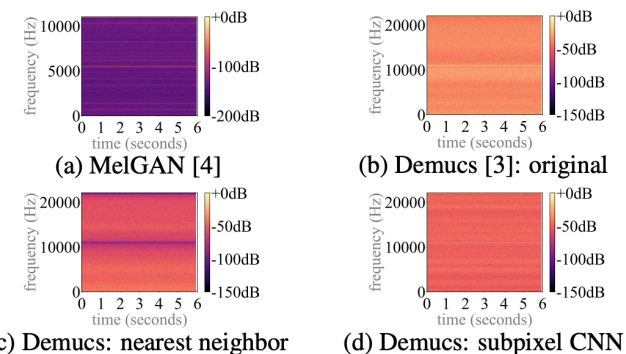
Jordi Pons, Santiago Pascual, Giulio Cengarle, Joan Serra

Dolby Laboratories

### ABSTRACT

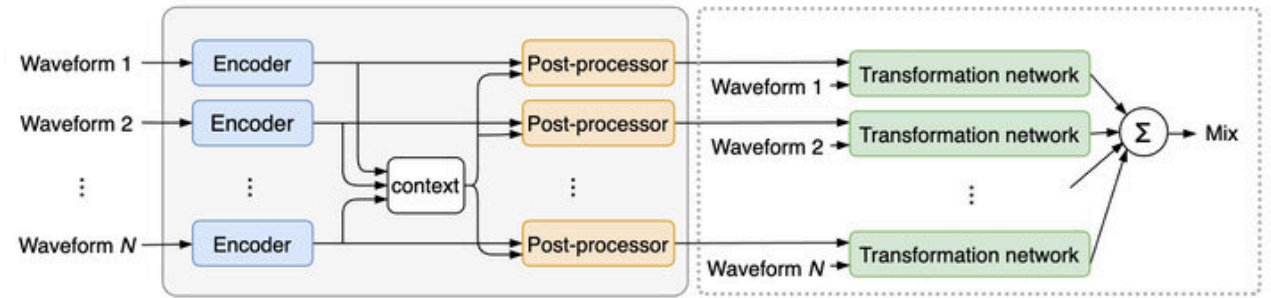
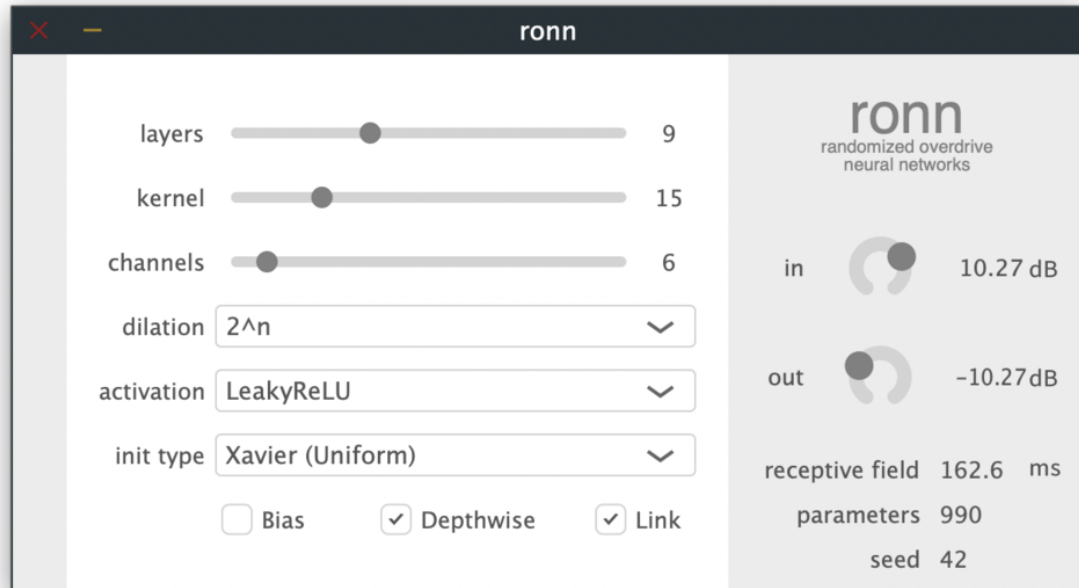
A number of recent advances in neural audio synthesis rely on up-sampling layers, which can introduce undesired artifacts. In computer vision, upsampling artifacts have been studied and are known as checkerboard artifacts (due to their characteristic visual pattern). However, their effect has been overlooked so far in audio processing. Here, we address this gap by studying this problem from the audio signal processing perspective. We first show that the main sources of upsampling artifacts are: (i) the tonal and filtering artifacts introduced by problematic upsampling operators, and (ii) the spectral replicas that emerge while upsampling. We then compare different upsampling layers, showing that nearest neighbor upsamplers can be an alternative to the problematic (but state-of-the-art) transposed and subpixel convolutions which are prone to introduce tonal artifacts.

**Index Terms** — upsampling, neural networks, synthesis, audio.



**Fig. 1. Upsampling artifacts after initialization:** tonal artifacts (horizontal lines: a,b,d) and filtering artifacts (horizontal valley: c). Input: white noise. MelGAN operates at 22kHz, Demucs at 44kHz.

# Neural audio effects



Randomized overdrive neural networks  
[arxiv.org/pdf/2005.00341.pdf](https://arxiv.org/pdf/2005.00341.pdf)

Differentiable console of neural audio effects  
[csteinmetz1.github.io/dmc-icassp2021/](https://csteinmetz1.github.io/dmc-icassp2021/)

Or audio style transfer effects by "Never Heard Before Sounds":  
[gan.style/poCw2CCrfzA?s=choral](https://gan.style/poCw2CCrfzA?s=choral)

—  
**Ethics.**

# Magenta on biases and power asymmetries

“While these choices (biases) are essential to make a model work, they bake in limitations by definition. It’s important to **make these choices clear and explicit in our musical context.**”

“Many AI music models embed structure from Western classical music (i.e. model bias). Melodies are divided into one of twelve pitches in an octave; like a piano’s black and white keys. In Thai classical music, the same octave is divided into seven pitches.”

“DDSP\* models thrive in **low-data environments typical of underrepresented music**. DDSP models are fast and lightweight, capable of rendering audio in-browser within seconds and with on-device privacy. Previous models were limited in scope to Western datasets, requiring hundreds of hours of recording data and powerful computing resources only available at places like Google.. ”

Stepping Towards Transcultural Machine Learning in Music ([magenta.tensorflow.org/transcultural](https://magenta.tensorflow.org/transcultural))

\*DDSP (differentiable digital signal processing) model was proposed by Magenta: [arxiv.org/abs/2001.04643](https://arxiv.org/abs/2001.04643)

# Magenta on cultural ownership

“Cultural ownership complicates this further; especially for underrepresented ones. We were worried non-Chinese listeners hearing the Guqin for the first time through our technology could form incorrect impressions of the instrument and, in turn, Chinese classical music.”

“DDSP \* is much like **a modern re-incarnation of sampling, a musical process which gave birth to new genres like hip hop**. But it could also set loose a new form of AI-powered cultural appropriation.”

“Is paramount to ensuring we do not appropriate, cause cultural harm, or implant users with wrong expectations through machine learning.”

Stepping Towards Transcultural Machine Learning in Music ([magenta.tensorflow.org/transcultural](https://magenta.tensorflow.org/transcultural))

\*DDSP (differentiable digital signal processing) model was proposed by Magenta: [arxiv.org/abs/2001.04643](https://arxiv.org/abs/2001.04643)



# On scraping and sharing data over the internet for music AI

**It's not ethical to scrape unlicensed copyrighted content.**

Note that, currently, large datasets are required for developing compelling generative models with machine learning.

For the music generation case, how much it costs to license 1M songs?

While it might be legal to share STFT transform of a song over the internet, is it ethical?

**Is scraping copyrighted media illegal in some countries?**

In the USA, it exists the Fair Use doctrine - allowing machine learning models to learn patterns from scraped data.

This legal case is always cited in this context: Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

# Kyle McDonald: on data biases

“Another significant challenge with data-driven algorithmic composition is: what data to use? Whose music counts?

When any automated creative system needs to be trained on a large number of cultural artifacts, **it can only perpetuate the dominance of what is already well-documented**. In music, this means a lot of Bach, Beethoven, and other old white European men. Two exceptions: some English and Irish folk music, and some video game music. The data is also selected by machine learning researchers, who are also a relatively homogenous group (though decreasingly so).”

[medium.com/artists-and-machine-intelligence/neural-nets-for-generating-music-f46dffa21c0](https://medium.com/artists-and-machine-intelligence/neural-nets-for-generating-music-f46dffa21c0)

# On impersonating or copying the style

**At music schools teach how to compose as Bach or Mozart as an exercise to learn composition.**

Tribute bands copy the style (including: outfit, attitude, and voice) of the original.

As far as I know, one does not need to ask for permission for replicating the composition style or singing (even with the exact same tone of voice).

**The red line is plagiarism.** In the case of a tribute band, you might be required to pay royalties in some cases.

**Music is like science, that "stands on the shoulders of giants".**

Copying or getting inspired by someone else should be legit.

**It is unclear whether it is unethical or not to copy the style of a specific artist.**

However, we need to be concerned by deep fakes.

**Is interesting to copy an already known art form or artist?**

# On copying the style: example

"Daddy's Car" is a song composed with Artificial Intelligence in the style of the Beatles.

**Is interesting to copy an already known art form or artist?**

Scientists at SONY have created the first-ever entire songs composed by AI: "Daddy's Car" and "Mister Shadow". The researchers have developed Flow Machines, a system that learns music styles and exploits unique combinations of style transfer, optimization, and interaction techniques, Flow Machines is meant to be an interactive tool that can suggest new melodic creative ideas. "Daddy's Car" is composed by Benoit Carré and François Pachet

Listen this song:

[youtube.com/watch?v=LSHZ\\_b05W7o](https://youtube.com/watch?v=LSHZ_b05W7o)

# Social implications

## **Music AI deep fakes, what are its social implications?**

Quality of the model is not yet there, what limits the social impact of those.

Deep fakes might obfuscate the AI-origin of the piece. Traceability could be key to limit the social impact of those.

This highlights the relevance of the licenses in machine learning.

What if one need to specify which model was used to create such piece?

## **Music AI can affect livelihoods of musicians and to the whole industry.**

The same way drum machines and sequencers affected drummers lives?

## **Are automatic music generation models a democratizing force for music production?**



# Ethics & aesthetics towards building an artistic discourse around music AI.

—

@jordiponsdotme - [www.jordipons.me](http://www.jordipons.me)

May 2021