



# AI Song Contest

## Human-AI co-creation in songwriting

Cheng-Zhi Anna Huang, Hendrik Vincent Kooops, Ed Newton-Rex, Monica Dinculescu, Carrie J. Cai

Motivation: to understand the gap between AI and musician needs

LSTM  
SampleRNN  
WaveNet  
Transformer  
VQ-VAE

Help! Not just any AI Help! You know I need a "darker" chorus Help!



# AI Song Contest

Organized by Karen van Dijk et al at VPRO  
Announced during ISMIR 2019

### Data

Conducted survey from 13 teams, 61 contestants

- How did teams decide **which aspects of song used AI vs composed by musicians?** What were the **trade-offs**?
- How did teams develop their AI system?
- How they incorporated their AI system into their **workflow** and **generated material** into their song?

### Approach

Qualitative analysis

- Musicians / developers needs
- Challenges and strategies for overcoming
- Design implications

## Three musician challenges

- 1 AI is not easily decomposable
- 2 AI is not context-aware
- 3 AI is not easily steerable

### 1 AI is not easily decomposable

End-to-end

I need to tweak the lyrics and melody!

**Dadabots x Portrait XO**

- Generated 10h of vocals in search of discernible lyrics and melody that goes well together
- The excerpts they found fueled backstory
- Artist composed duet by riffing along

### Musicians use modular musical building blocks

Music building blocks	Models & techniques
Lyrics	GPT2, LSTM, Transformer
Melody	CharRNN, SampleRNN, LSTM + CNN, WaveNet + LSTM, GAN, Markov model
Harmony	LSTM, RNN autoencoder, GAN, Markov model
Bassline	LSTM + CNN, WaveNet + LSTM, GAN
Drums	DrumRNN, Neural Drum Machine, SampleRNN, Markov model
Multi-part	MusicVAE trio (melody, bass, drums), MiniVAE trio, Coconet/Coucou (4-part counterpoint), MusicAutobot (melody, accompaniment), Transformer (full arrangement)
Structure	Markov model
Vocal synthesis	WaveNet, SampleRNN, Vocaloid, Sinsy, Mellotron, Emvoice, Vocaloid, custom vocal assistant
Instrument synthesis	SampleRNN, WaveGAN, DDSF

Table 1. Overview of musical building blocks used by teams.

### Musicians strategies in combining building blocks

**manually stitch**

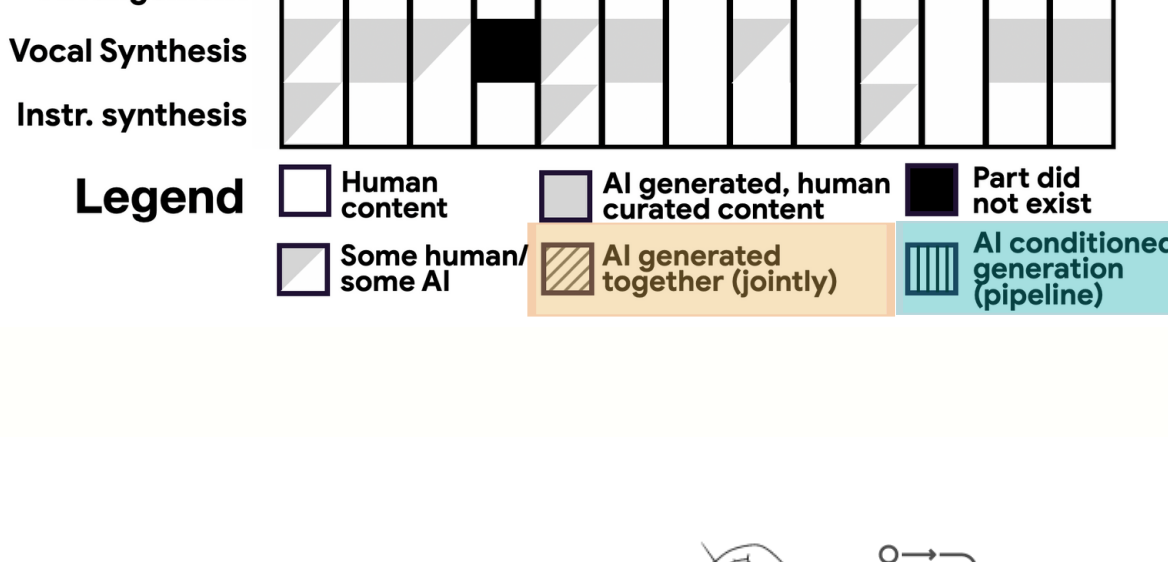
i.e. generate melodic lines, manually stack to create harmony

**pipeline (conditioned)**

i.e. generate chords, then melodies and bass lines separately

**multiple-parts together (jointly)**

i.e. using MusicVAE, Coconet, MusicAutobot, Transformer



### Why teams used one strategy vs another?

- Teams w/ professional musicians used AI to generate lyrics and melodic lines to leave create space to musicians, or lead sheets
- Teams w/ more ML and less musical expertise used ML that jointly generates multiple parts, to have larger coherent building blocks as a starting point

### 2 AI not context aware

Here's some lyrics, more lyrics

Which of them fit together?

**Uncanny Valley**

Used stress patterns to match lyrics and melody algorithmically

### Generate and curate

- Offline batch mode**  
Generate many at once
- Online active co-creation**  
"Jam" with ML. Feedback loop: "prime" ML, listen to its outputs, modify the input sequence to steer ML towards desired musical results

"Rejection sampling"  
Manually cherry pick. Find ones that fit with the musical context. Or train a "hit melody" classifier

### 3 AI not easily steerable

I need the chorus to sound "darker", also a cadence like "Bach"

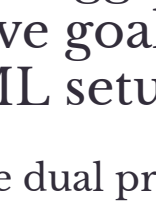
**Dadabots x Portrait XO**

- Tried out 3 models for rewriting and reharmonization to achieved desired musical effect
- Used 7 different models total

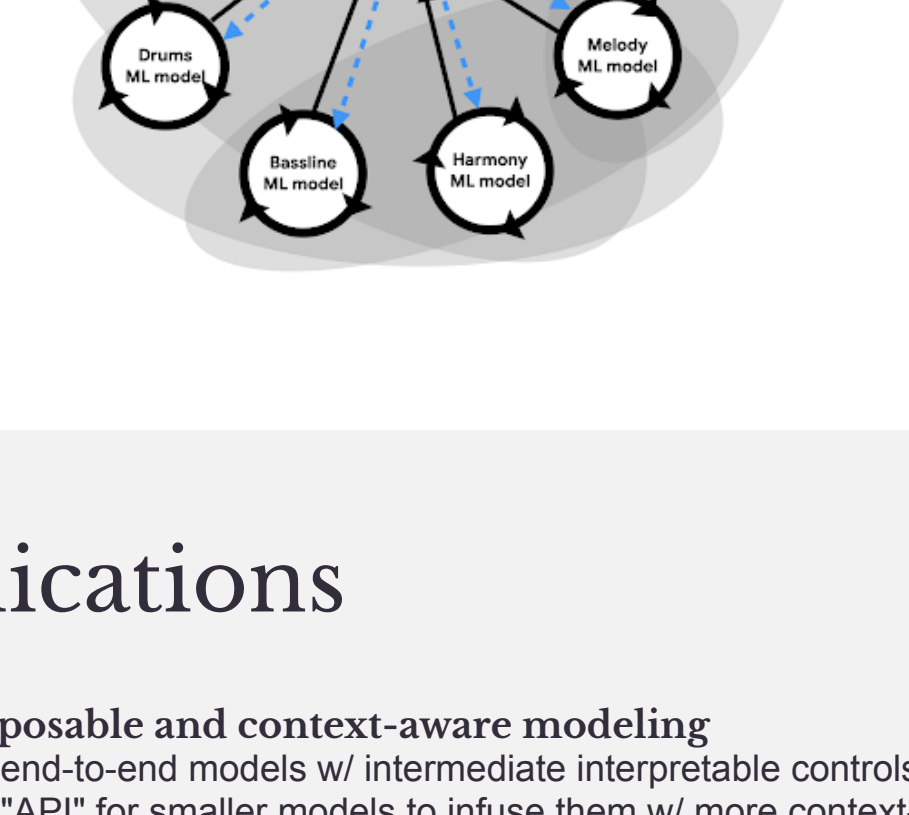
In summary

## Three musician challenges

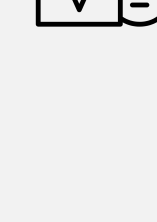
- 1 AI is not easily decomposable
- 2 AI is not context-aware
- 3 AI is not easily steerable



Interleave debugging creative goal with ML setup  
Juggle dual process



## Design Implications



- **De-composable and context-aware modeling**
- Design end-to-end models w/ intermediate interpretable controls
- Design "API" for smaller models to infuse them w/ more context-awareness and user-facing controls
- Hybrid: combine global context and flexibility



- **Musician defined (vs AI-defined) building blocks**
- How researchers decompose music impact how musicians think about music, or who can benefit from these tools
- Allow musicians define first-class building blocks and principles
- Design for musicians' workflows

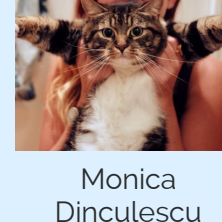
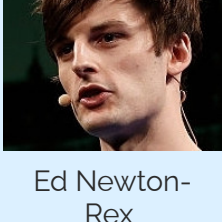
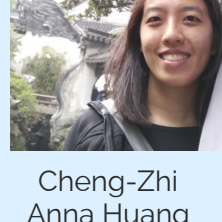
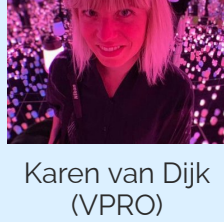


- **Expose ML controls directly in UI**
- Support ML and musical exploration simultaneously (i.e. Magenta Studio: multiple ML models as plugins in Ableton Live)
- Larger musical context potentially available to ML models
- Allow users to semantically steer ML
- Scaffold strategic parts of model exploration and selection (i.e. suggest model combinations, workflow heuristics)

BIG THANKS to AI Song Contest organizer

SPECIAL THANKS to the 13 teams from the AI Song Contest teams for their amazing contributions that made this research possible!

Paper brought to you by



Karen van Dijk (VPRO)

Cheng-Zhi Anna Huang (Google)

Hendrik Vincent Kooops (RTL Netherlands)

Ed Newton-Rex (ByteDance)

Monica Dinculescu (Google)

Carrie J. Cai (Google)