

# Generative AI helps one express things for which they may not have expressions (yet)

Bob L. T. Sturm

bobs@kth.se

Tal, Musik och Hörsel, School of Electrical Engineering and Computer Science, KTH Royal Institute of Engineering  
Stockholm, Sweden



Figure 1: This is a portion of a digital image generated by the generative AI system VQGAN-CLIP [1, 2] in response to the text prompt, “Who removed the light from the sky?” This appears to be a sky with a burning patch where the sun once floated.

## ABSTRACT

I describe how several generative AI systems have helped me express things for which I did not yet have expressions, or did not know that I wanted to or should express. This partnership has resulted in several multimedia art pieces, and my experience in their creation shed light on several questions relevant to this workshop.

## CCS CONCEPTS

- Computer systems organization → Embedded systems; Redundancy; Robotics;
- Networks → Network reliability.

## KEYWORDS

computational creativity, co-creation, machine learning, multimedia

### ACM Reference Format:

Bob L. T. Sturm. 2018. Generative AI helps one express things for which they may not have expressions (yet). In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX'). ACM, New York, NY, USA, 5 pages. <https://doi.org/XXXXXXX.XXXXXXX>*

Unpublished working draft. Not for distribution.  
Permission to make digital or hard copies of all or part of this work for personal or  
classroom use is granted without fee provided that copies are not made or distributed  
for profit or commercial advantage and that copies bear this notice and the full citation  
on the first page. Copyrights for components of this work owned by others than ACM  
must be honored. Abstracting with credit is permitted. To copy otherwise, or republish,  
to post on servers or to redistribute to lists, requires prior specific permission and/or a  
fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/XXXXXXX.XXXXXXX>

2022-02-18 18:58. Page 1 of 1–5.

## 1 INTRODUCTION

This paper documents four multimedia pieces resulting from my partnership with several generative AI systems:

- (1) *Evigt Förlorad (Forever Lost)* (3:51)<sup>1</sup>
- (2) *They are everywhere* (3:44)<sup>2</sup>
- (3) *A slow waltz* (3:46)<sup>3</sup>
- (4) *The shadow still lay where he had been standing* (3:00)<sup>4</sup>

The systems I use to generate musical, textual and visual material are *folk-rnn* [4, 6], Music Transformer [5], GPT-J-6B,<sup>5</sup> and VQGAN-CLIP<sup>6</sup> [1, 2]. I use *folk-rnn* to generate melodic material, sometimes resembling tunes in their training datasets of traditional Irish or Swedish dance music. The Music Transformer generates harmonic piano accompaniment for given melodies, which often can be used as is, or scored for other instruments. The generative text model GPT-J-6B is trained on an 800+ GB collection of English text [3], and generates text in response to a given prompt, e.g., “Write a poem.” Finally, the VQGAN-CLIP system generates digital images given a textual prompt, e.g., “Who removed the light from the sky?” as seen in Fig. 1.

As a composer I notice several benefits in my creative partnerships with a variety of generative AI systems. These systems are always on and ready to respond to input. They do not require positive reinforcement and are indifferent to personality. For instance, I can take an output of a system and criticise it, modify it, or trash

<sup>1</sup><https://www.aisongcontest.com/participants/sthlm-2021>

<sup>2</sup><https://bit.ly/3uvCB41>

<sup>3</sup><https://bit.ly/3J7smC>

<sup>4</sup><https://bit.ly/3GuSe9t>

<sup>5</sup><https://github.com/kingoflolz/mesh-transformer-jax/#gpt-j-6b>

<sup>6</sup><https://github.com/nerdyrodent/VQGAN-CLIP>

The figure consists of two musical staves. The top staff (lines 117-124) shows a melody in G major with a 2/4 time signature. The bottom staff (lines 125-137) shows a modified melody in G major with a 2/4 time signature. The melody is labeled C(AABBC)2CA. The bottom staff is divided into three sections: A (lines 126-129), B (lines 130-132), and C (lines 133-137).

**Figure 2: Top: melody generated by *folk-rnn* trained on Irish traditional dance music. Bottom: modified melody featured in *Evigt Förlorad (Forever Lost)*.**

it, without feeling guilt. I can use these systems to off-load particular tasks without the baggage of working with other personalities and abilities. While these systems sometimes act in unpredictable ways that can be hard to control, this can produce opportunities for creativity and suggest alternatives. Even when they fail they can become useful. They can also act as gateways to the surreal, which arises from learning tasks that transcend operationalization and quantitative evaluation.

## 2 EXEMPLARY PIECES

The following four pieces feature materials produced by several generative AI systems. I describe the background and my process for each piece, and how generative AI was an essential partner.

### 2.1 “Evigt Förlorad (Forever Lost)”

The melody of this song comes from a tune generated by a *folk-rnn* system. The lyrics are a Swedish translation of text I created through interacting with GPT-J-6B. It is about real grief over a six-month-old puppy being struck and killed by a car that never stopped.<sup>7</sup>

This song was written at a time when I was experimenting with ideas for The AI Song Contest 2021.<sup>8</sup> Around that time I happened across the melody at the top of Fig. 2, generated by *folk-rnn* but rejected by an “artificial critic” I engineered to select tunes that fit a specific form of dance music [7]. I found the tune while checking the rejections to see how well the critic was working, and set it

<sup>7</sup>In 2020 my wife and I adopted a puppy. We spent two wonderful months growing together until the puppy died on her six-month birthday after escaping from her harness and running into a busy road while friends took her on a walk.

<sup>8</sup><https://www.aisongcontest.com>

aside because I found it lyrical. I proposed to my team that we use this for part of our entry to the contest, but I did not know what it could be about. Then I thought about that tragic experience from a few months prior.

The greatest extent to which I could emotionally engage with this loss was prompting GPT-J-6B with the situation, and then curating lines that expressed how I felt and that fit the melody. I gave the prompt:

Our girl was only six months old  
The car never stopped  
and GPT-J-6B responded  
On the way home it flew  
An idiot flew my girl the fuck away  
She was only six months old  
An idiot flew my girl the fuck away  
On the way home it flew  
This is the story of a little baby who just wants to get home.

Another iteration gave:

She said 'I'm okay, I'm okay'  
She was a girl we saw  
Car for sale by the side of the road  
We never knew what it would take to raise her  
Loved her in the front seat we used to love her  
I remember when she would pick our baby up and give it her  
Or she would pick the little one up  
And roll us all out to church  
But she only loved us in the back seat

And another iteration:

When we went out to eat the waiter threw her in the car  
Got in the driver seat and drove away  
Dentist – or just a good entertainer – never found her.  
So she's gone, never to be found  
Walking down the streets looking for me  
I look around and there's nobody to see her  
Choke on her pink and blue and throw her in a bucket  
The grave of her heart I've dug  
In my backyard

My wife and I curated from these lines, added a few special lines of our own, and had the lyrics of the song:

A  
On the way home, our baby girl flew  
So she's gone, never to be found, our baby girl flew.  
What are we to do? This world that we've built?  
This child, our angel, our baby girl,  
the world that we've built?

B1  
She was only six months old  
An idiot flew our little girl the fuck away  
So she is gone, never to be found  
An idiot flew our little girl the fuck away

C  
Never stopped, the car never stopped  
We will never know what it'll take to raise her.

233

234 B2

235 She had just learned to run like the wind, at six months old,  
236 But an idiot flew our little girl the fuck away.  
237 She loved to sit in the sun with her toys, at six months old,  
238 But an idiot flew our little girl the fuck away.

239  
240 A team member translated the lyrics to Swedish,

241 The final melody appears at bottom of Fig. 2 and follows the  
242 structure C(AABBC)2CA. I composed the piano accompaniment  
243 from material generated by Music Transformer given the *folk-rnn*  
244 melody. The recording features this accompaniment played by a  
245 Disklavier mechanical piano, along with voice, accordion and fiddle.

246 Generative AI provided me a path through a dark personal chapter  
247 that would have required too much courage to do alone. I did  
248 not know I needed to write this song, and never have I thought I  
249 would write something so personal. A solo accordion performance  
250 with pictures of our puppy is here: <https://youtu.be/rRu8LAhEp-w>.

251

## 252 “They are everywhere”

253

254 A folk-rnn model trained on Irish traditional dance music generated  
255 the tune seen at top in Fig. 3. After working with it on my accordion,  
256 I changed it to the version seen at bottom in Fig. 3, and felt it best  
257 played as a slow instrumental. To find a topic for this work, I queried  
258 GPT-J-6B with the prompt, “It used to be me.” and it generated:

259

260 The little girl who read only at bedtime, who read the  
261 books I was supposed to read, under the covers, be-  
262 cause they were too hard to concentrate on during the  
263 day. I liked the stories they told about fairies, pirates,  
264 and knights. I liked to imagine I lived in their world,  
265 one where bad people died horribly, one where good  
266 people saved each other, one where all the boys were  
267 brave and handsome, and where all the girls were  
268 pretty and sweet. It was a world of magic.

269

270 Now it was a world where everything was broken.  
271 The monsters still existed, but they were everywhere.  
272 They crawled in your veins, and the things you found  
273 in the shadows of your mind. They were in the blood  
274 you drew from a vein on your arm, in the saliva that  
275 dribbled from your lips. They moved in your brain,  
276 and they moved in the air, and they moved in the  
277 shadows of the world, and they were everywhere.  
278 They were everywhere,

279

280 This reminded me of my wife, who as a child would read under  
281 the covers with a flashlight after her parents put her to bed. I created  
282 a new prompt “It used to be me. The little girl who reads at bedtime,  
283 under the covers with a flashlight, against my parents’ wishes.”  
284 GPT-J-6B responded:

285

286 She was not allowed out of the house, in case she was  
287 picked up and sent to a children’s home or worse,  
288 because that’s how I was raised. But a lot has changed  
289 in my life. Now I have a son, who loves books. And he  
290 reads anything he wants. I love being able to read to  
291 him. It’s kind of amazing. This is how things should  
292 be.



Figure 3: Top: melody generated by *folk-rnn* trained on Irish traditional dance music. Bottom: modified melody featured in *They are everywhere*.

I iterated with the output of GPT-J-6B, building up prompts for generating more text, and finally formed six stanzas of free verse:

I used to be me.

The little girl who reads at bedtime,  
under the covers with a flashlight,  
against my parents’ wishes.

I liked the stories about fairies, pirates,  
monsters and knights.

I imagined I lived in their world,  
one where good people save each other,  
and where bad people die horribly.

Now I live in a world where everything is broken,  
and there are monsters everywhere.

They crawl in my veins, in the shadows of my mind.  
They are everywhere.

Now I have a child, who loves books.  
A little girl who reads at bedtime,  
under the covers with a flashlight,  
against my own wishes.

She likes the stories about fairies, pirates,  
monsters and knights.

She imagines she lives in their world,  
one where good people save each other,  
and where bad people die horribly.

But, she lives in a world where everything is broken,  
and there are monsters everywhere.

They crawl in her veins, in the shadows of her mind.  
They are everywhere.

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348



**Figure 4:** Top: melody generated by folk-rnn trained on Scandinavian traditional dance music. Bottom: modified melody featured in *A slow waltz*.

Rather than a song, I imagined accompanying my wife reading the text to form a dramatic experience. My wife’s voice is processed with various effects that grow in prominence as the piece proceeds. I created video imagery from stock videos, and found footage from an archivist in San Diego.<sup>9</sup>

### 2.3 “A slow waltz”

This piece is arranged around a melody generated by folk-rnn trained on Swedish traditional dance music [6], shown in Fig. 4 in the original and modified forms. Much in the same way as for *They are everywhere*, I worked with GPT-J-6B to create a script around which I would play the melody. The end result is a slow waltz of a conversation between two faceless characters, one of whom bemoans the accordion and praises the saxophone, and eventually tries their hand at playing the former.

I created visual material for this piece by prompting VQGAN-CLIP with text drawn from the script. Figure 5 shows a screen grab generated from the prompt She said, Oh great, more accordion? | stipple. This system iteratively perturbs a random image in ways that increase the likelihood of being captioned with the given prompt, as well looking like a real image I use the updated image of each iteration as a video frame, and then stitch these together to accompany the dialogue.

### 2.4 “The shadow still lay where he had been standing”

I envision this work as the first “episode” of a long-form composition exploring surrealist notions of GPT-J-6B and VQGAN-CLIP. As in *A slow waltz* and *They are everywhere*, I built up a short script with GPT-J-6B, first by prompting “deep and moving (a poem)”. It responded, The shadow still lay where he had been standing and some other stuff. After going back and forth with GPT, I arrived



**Figure 5:** Still from the video accompanying *A slow waltz*, created by VQGAN-CLIP from the prompt, She said, Oh great, more accordion? | stipple.

to a brief text that I liked. At the beginning of the piece, we hear a voice saying:

The shadow still lay where he had been standing.  
And then it too was gone.  
That was the last I saw of him.  
The last I saw of anything.  
Who came and took the light from the sky?

We then experience visions of other worlds and strange figures. One particularly beautiful scene is created by VQGAN-CLIP from the prompt Who came and took the light from the sky? | cryengine, seen in Fig. ???. The system created what looks like a sunset sky with a small dancing flame where the sun had been.

After a while we hear the voice say, “Maybe I should call The Electrician.” “The Electrician” is then revealed, seen in Fig. 6, hallucinated by VQGAN-CLIP from the prompt He holds the sun in his hands | cryengine. Who does this voice belong to? Why do they suddenly have a lack of vision? Whose shadow was that? Who is The Electrician? These will be answered in future episodes, but so far derived from interacting with GPT-J-6B and VQGAN-CLIP: the one we hear speaking is Arthur, an Android who just joined a multilevel marketing scheme selling what he has been told are “power poses”, which he misheard as “power poésie” (he fancies himself a poet would like to make a living from it).

Melodic material for this episode was generated by *folk-rnn* trained on Irish traditional music, shown in original and altered forms in Fig. 7. I created a MIDI-file from several repetitions and variations of the melody and then queried Music Transformer to create an accompaniment. I found curious results after setting the tempo to be very high, and added a brief introduction setting the rhythmic feel. The failure of a notation program in converting the generated MIDI data, combined with decreasing the tempo, results in interesting rhythmic characteristics.<sup>10</sup> I took one of these outcomes and scored it for an ensemble of percussion, marimba, harp, celesta, piano, and double bass.

<sup>9</sup><https://www.youtube.com/c/peahix/videos>

<sup>10</sup>Intermediate musical examples can be heard here: <https://bit.ly/3GS3sVQ>

465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478



479 **Figure 6: Still from the video accompanying *The shadow still***  
480 ***lay where he had been standing*, created by VQGAN-CLIP**  
481 **from the prompt, He holds the sun in his hands |**  
482 **cryengine. (Text has been overlaid.)**

483  
484  
485  
486  
487  
488  
489  
490

491  
492  
493  
494  
495  
496  
497  
498



499 **Figure 7: Top: melody generated by folk-rnn trained on Irish**  
500 **traditional dance music. Bottom: modified melody featured**  
501 **in *The shadow still lay where he had been standing*.**

502  
503

### 3 DISCUSSION

506  
507  
508  
509  
510  
511  
512  
513  
514  
515

*Evigt Förlorad (Forever Lost), They are everywhere* and *A slow waltz* all served as experiments in ways of working with these tools. It was important for me to document how particular interactions occurred, and where creative possibilities arose. In my growing composition, *The shadow still lay where he had been standing*, however, tracking provenance becomes an impediment to creation. Assigning credit to parts of pieces is like recognizing a particular hair on a paintbrush for a specific stroke. It is interesting academically speaking, but is not primary to what I am trying to accomplish: express things for which I do not yet have expressions.

516  
517  
518  
519  
520  
521  
522

Discuss other things here. Does generative AI go beyond intelligent interaction? How do you design in this space? What design patterns do we need to think about? Failure, abuse, exploration and exploitation, rapid feedback and turnaround. Serendipity, uncertainty and randomness. Is presenting AI as a desirable and “objective” method appropriate for generative AI? “The ability of a machine to do or outdo something humans do is interesting once at

most” (Jan Swafford "The intelligence of Bodies: The philosophical and musical failings of “Beethoven X: The AI Project” 2021)

523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580

### ACKNOWLEDGMENTS

This work was supported in part by the project ERC-2019-COG No. 864189 MUSAiC: Music at the Frontiers of Artificial Creativity and Criticism.

### REFERENCES

- [1] Jong Wook Kim Gretchen Krueger Sandhini Agarwal Alec Radford, Ilya Sutskever. 2021. CLIP: Connecting Text and Images. <https://openai.com/blog/clip/>
- [2] Patrick Esser, Robin Rombach, and Björn Ommer. 2020. Taming Transformers for High-Resolution Image Synthesis. *arXiv* (2020). arXiv:2012.09841 [cs.CV]
- [3] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, Shawn Presser, and Connor Leahy. 2021. The Pile: An 800GB Dataset of Diverse Text for Language Modeling. *arXiv* abs/2101.00027 (2021).
- [4] E. Hallström, S. Mossmyr, B. L. Sturm, V. H. Vegeborn, and J. Wedin. 2019. From Jigs and Reels to Schottisar och Polskor: Generating Scandinavian-like Folk Music with Deep Recurrent Networks. In *Proc. Sound and Music Computing Conf.*
- [5] Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Curtis Hawthorne, Andrew M Dai, Matthew D Hoffman, and Douglas Eck. 2018. Music Transformer: Generating Music with Long-Term Structure. *arXiv preprint arXiv:1809.04281* (2018).
- [6] B. L. Sturm, J. F. Santos, O. Ben-Tal, and I. Korshunova. 2016. Music Transcription Modelling and Composition Using Deep Learning. In *Proc. Conf. Computer Simulation of Musical Creativity*. Huddersfield, UK.
- [7] B. L. T. Sturm. 2021. An Artificial Critic of Irish Double Jigs. In *Proc. AI Music Creativity*.