How Music AI Is Useful
Engagements with Composers, Performers and Audiences

ODED BEN-TAL, MATTHEW TOBIAS HARRIS AND BOB L.T. STURM

Critical but often overlooked research questions in artificial intelligence applied to music involve the impact of the results for music. How and to what extent does such research contribute to the domain of music? How are the resulting models useful for music practitioners? This article describes work arising from research engaging with composers, musicians and audiences to address such questions: two websites that make their AI models accessible to a wide audience and a professionally recorded album released to expert reviewers to gauge the plausibility of AI-generated material. The authors describe the use of their models as tools for cocreation. Evaluating AI research and music models in such ways illuminates their impact on music-making.

When applying artificial intelligence (AI) or, more specifically machine-learning techniques that underpin current research, in creative fields, one critical question to answer is: Why should such technology be applied to such an activity? Much research in this area tries to answer different questions, such as: Can AI system XYZ fool humans into believing its creations are by humans [1]? Or: How statistically similar are AI-generated outputs and the dataset used for training [2]? Applied to art, these methods lack meaning. Evaluating creative systems requires looking beyond the generated outputs to the role of expertise and the perspectives of different target audiences [3].

Motivated by Wagstaff [4], our research addresses the application of AI to musical domains. Our aims are to test how such AI systems can operate as part of a music ecosystem and to engage practitioners with the questions, problems, opportunities and challenges that AI raises for music. We engage a range of practitioners with AI models we have developed and critically examine ways these models are used within musical practices. This highlights the contribution AI technology can make to music and suggests where future developments could be most fruitful.

Machine learning (ML) involves making a computer learn patterns by example, making it attractive for modeling, generating and transforming music [5–10]. Most current work in music ML revolves around tasks that have been explored computationally almost as long as computers have existed [11], e.g. melody and harmony generation in a few styles. This reflects both the availability of data needed to train the models and the extensive theorizing that make possible interpretation of the outputs.

Our research [12] applies deep ML to specific domains of Western European folk music, creating models of a type we call “folk-rnn.” One data source consists of text-based transcriptions (ABC notation [13]) of traditional dance music mostly from Ireland and the United Kingdom, crowdsourced at The Session website [14]. After extensive data-cleaning—removing incomplete transcriptions, comments, chord progressions and unrelated examples such as John Cage’s 4’33”—our dataset consists of over 23,000 transcriptions. Another source is Scandinavian folk music transcriptions expressed in the same vocabulary [15]. Our models extract from these datasets probability distributions that can be used to generate new transcriptions [16].

We have used several methods to explore potential contributions of such ML models to music [17]: We have performed musical analysis on material generated by a model [18], examined the system’s performance when seeded with out-of-sample material [19], solicited opinions from users at The Session [20] and used the system for composition [21]. Here, we extend this evaluation further by engaging wider audiences with the models and their generated music.

ACCESSIBLE ONLINE IMPLEMENTATIONS
We created two websites around our models to make them accessible. One is an interactive interface and the other is a growing repository of music generated by or with the first [22].
Two Web Resources: www.folkrnn.org and www.themachinefolksession.org

The folk-rnn website presents an optimized, server-based implementation of our models and a user interface that exposes functionality in a more straightforward and appealing manner than the command-line. The interface comprises a panel (Fig. 1) that presents music generation controls and a main section that scrolls to hold each tune as a user generates it. On the initial page load, this section shows site information including a video demonstration.

Clicking the Compose button, the most prominent control on the page, results in a new tune appearing character by character as it generates. Further controls provide for iterative or deliberate use:

- **Model** differentiates the data used for training.
- **Temperature**, raised or lowered, determines how “adventurously” the model acts.
- **Seed** controls the internal pseudorandom state to produce new transcriptions for the same parameters. It will change for each tune generated unless “pinned” by manual input.
- **Meter** and **Mode** can be selected from a set of options, e.g. 4/4 or 6/8, C major or C dorian.
- **Initial ABC** text box allows a user to specify the beginning of a tune, which the model then completes.

In addition to the textual ABC [23] representation output, staff notation and audio playback are provided; playback animation links all three representations. A user can download the result in MIDI format or archive the result.

The themachinefolksession website serves as a community archive for music created by or with folk-rnn models, primarily organized around tunes. Figure 2 shows two user-submitted tunes—one is clearly outside the idiom of Irish traditional music. On any tune's page, one can see the original submission along with any backstory, settings (e.g. an edit or variation of the original) and performances (video or audio recordings), comments and links to related events. Registered users can add tunes to personal “tunebooks.” Inspired by folk sessions elsewhere, we experimented with features such as “tune of the month” in which each member contributes their own take on a particular tune (this was not a success).

**Usage**

Our analysis of the sites’ server data shows their use and the impact of media attention on our research. During the first 235 days of activity at folkrnn.org, 24,462 tunes were generated by approximately 5,700 users (an order of magnitude more than the approximately 250 people who engaged with a command-line tool in its first three years of existence) [24]. Activity for the first 18 weeks averages a median of 155 tunes weekly. Since then, overall use has increased, with a median of 665 tunes generated weekly (as of August 2019). This period also features usage spikes. The largest, correlating with a German media report [25], shows an 18.4× increase in weekly tunes generated.

This data provides evidence of human-machine cocomposition using folkrnn.org. There are 4,007 transcriptions for which one or more parameters have been changed prior to generation. We see an average of 6 (mean: 5.9, stddev: 8.7) iterations in such processes, accounting for 57% of all transcriptions generated. **Temperature** is the most-used parameter at 40%. This has the simplest action of the UI parameters—a simple numeric value. Changing this parameter can also result in dramatic changes in generated material; increasing the temperature from 1 to 2 will often yield tunes that do not sound traditional (see Fig. 2: “Stockhausen's Polka”). **Initial ABC** is used 20% of the time, which is notable as this requires text manipulation on the user's part.

The strongest metric of cocomposition available on folkrnn.org is whether **Initial ABC** contains a fragment of the previous generated transcription. This suggests the user has identified an interesting or useful portion and wishes to seed the next generation with it—in essence, performing “autocomplete.” Testing for sequences of five or more char-
acters (e.g. five pitches), we find this happened 283 times, or 2% of the time.

We find that 239 of the “iterative” folkrnn.org transcriptions have been archived to themachinefolksession, such as the user-named “The Green Electrodes” [26]. This was generated by a user on folkrnn.org in the key of C dorian. The user then submitted a “setting” that transposed it to the key E dorian but otherwise left it unchanged. This shows one limitation of folkrnn.org: All transcriptions are generated in a mode built on the first pitch of C (a consequence of an optimization made while training the model). This also shows that users have employed the manual editing features of themachinefolksession to work around this limitation.

The archived tune “Rounding Derry” [27] shows direct evidence of user intent: The user generated “FolkRNN Tune 24807” on a fresh load of folkrnn.org, i.e. using default parameters and a randomized seed. The user played this tune twice, then selected the ABC phrase “CAGECG/EFCG FDBG” and entered this as Initial ABC. The user generated the next iteration, played it back, named it and archived the result on themachinefolksession, writing, “Generated from a pleasant 2 measure section of a random sequence, I liked this particularly because of the first 4 bars and then the jump to the 10th interval key center (?!) in the second section.”

During the first 235 days of themachinefolksession site activity, 551 tunes were archived, of which 80% were generated at folkrnn.org. Of these 551, 15% have had one or more further iterations contributed. The two websites continue to document human-machine cocreations. As of 28 May 2021, themachinefolksession hosts 71 recordings and 1010 transcriptions; folkrnn.org has generated a total of 65,912 transcriptions.

\[\text{Fig. 2. Two tunes generated on folkrnn.org and archived at themachinefolksession. [© Oded Ben-Tal. User-generated content published under Creative Commons BY-SA Attribution-ShareAlike license.]}\]

result on themachinefolksession, writing, “Generated from a pleasant 2 measure section of a random sequence, I liked this particularly because of the first 4 bars and then the jump to the 10th interval key center (?) in the second section.”

During the first 235 days of themachinefolksession site activity, 551 tunes were archived, of which 80% were generated at folkrnn.org. Of these 551, 15% have had one or more further iterations contributed. The two websites continue to document human-machine cocreations. As of 28 May 2021, themachinefolksession hosts 71 recordings and 1010 transcriptions; folkrnn.org has generated a total of 65,912 transcriptions.

\[\text{LET’S HAVE ANOTHER GAN AIM: AN EXPERIMENTAL TRADITIONAL ALBUM}\]

In January 2018, we recorded a 45-minute album [28] with professional musicians at the Visconti studio, Kingston University, aiming to make an album that could be considered successful as Irish traditional music. We hired Daren Banarse [29], a musician we have worked with previously, to perform
AI-generated material in real musical contexts. The symbolic representation used by our models only provides the “bones” and does not explicitly specify critical nuances of Irish traditional music. It is thus necessary to have experienced musicians to render the material plausibly. This album was an extension of our experience with the musicians in concerts and was aimed at reaching wider dissemination within a context relevant to the training data’s specific domain.

*Let’s Have Another Gan Ainm* contains 31 tunes, 20 of which come from material generated by folk-rnn. The Gaelic phrase *gan ainm* means “without title”; each *gan ainm* on the album comes from generated material. Banarsë curated this material from 100,000 transcriptions, which we have assembled in 34 volumes [30]. He chose material from only six of those volumes and made changes to each tune. Although they tend to be small edits [31], some are musically significant. One major reason Banarsë cited was to improve musical flow. Many changes are at “linking points”—adding first and second endings to enable linking backward for repeats and forward to the second phrase, or changing the end of a tune for smoother transition to the next one. He also corrected some “mistakes,” e.g. a few bars with missing eighth notes (a human-like error occurring in the training data). Another aspect of Banarsë’s editing is the balance between conformity to common patterns and the inclusion of distinctive features. In some instances he reinforced repetition to improve structure (e.g. in the B part of tune #2375, the second *gan ainm* in the first track). In others he changed notes when he deemed a tune too mundane. Figure 3 (tune #6582, the third *gan ainm* in track three) shows a transcription generated by a folk-rnn model and the changes Banarsë made for the album. All changes are shown in the technical report [32].

Banarsë explains the changes to the opening:

Bars 1, 2 and 3 are each made up of a mini call and response—2 beats call, 2 beats response. I thought the 3rd response was too similar to bar 2, starting on a A, and not really seeing anything interesting. My rewritten response provides a mix between an inversion of the call, and a more interesting end to the 4 bar phrase [33].]

Some additional changes happened in the recording session itself when a musician’s variant was judged by the others to be better than the notated version (e.g. the end of the B part in the first *gan ainm* in track 3).

We privately released *Let’s Have Another Gan Ainm* online in March 2018 with the following information on the webpage:

During the Summer of 2017, three generations of the Ó Conaill family gathered at the family home in Roscom-
common to celebrate the life and legacy of Dónal Ó Conaill. The late father and grandfather to the Ó Conaill family, Dónal was quietly dedicated to the tradition, and known for collecting local tunes without names which he passed on to his family. His daughters, Caitlin and Úna, are joined by their children and family friends to make a recording of the best of these tunes, along with some of Dónal’s personal favourites.

We disguised the computer’s role to garner reactions to the album and not the technology [34], to prevent bias, e.g. when a result sounds better than a listener thought possible for a machine or when a listener is prejudiced against machine-created music. The latter is clearly evinced by comments on a Daily Mail article about our work [35]. The journalist embedded a 30-second excerpt of supposedly computer-generated traditional music. Reader comments ranged from negative (“Until they find a way to inject heart and soul into a computer it won’t happen. “Totally lifeless without warmth.”) to hostile (“Stupid idea, stupid outcome.” “This computerized ‘AI’ is just so non musically untalented lazy nerds can infiltrate the world of true musicians who love, created, and write the music from the joy, hurt, and life emanating from their hearts.”). In actuality, the journalist had accidentally excerpted a real tune, but many commenters heard a “robotic Irish jig.”

Reviewers of our album were positive and clearly heard the music as sitting comfortably within the tradition from which the training data comes. Referring to the backstory we posted, one reviewer wrote “Caitlin and Úna Ó Conaill . . . have done lovers of Irish traditional music an immense favour by allowing us this snapshot of a family reuniting to make delightful music” [36].

When we revealed the album’s true nature to reviewers, not one reacted negatively. We received by email interesting comments from one expert, Kevin McDermott. He still considered most of the tunes believable—some very successful, though two were odd or failures. He related specific tunes to musical subdomains. Referring to track 6, he wrote, “the ascent to the high note in the turn sounds like stuff young composers like the lads in Socks In The Frying Pan are writing”; on track 10, “the second [gan ainm in the set] is spot-on: a fine traditional jig which bears all the hallmarks of one from the late 18 to the mid-19C” [37].

We see that while folk-rnn models are generally successful for the style, they could be improved by better handling of particular local context such as the musical meaning of 1st and 2nd ending and helping human users find outputs suitable for their needs. We can also see that experts make finer distinctions about different parts of this musical corpus. A data collection of 23,000 examples is rather large compared to similar research in AI (e.g. building a model on Bach’s 371 chorale harmonizations). Larger amounts of data assist the model in imitation. At the same time, perhaps nuances are lost by aggregating the transcriptions without regard to how they belong to subclasses, e.g. dance types.

GOING BEYOND THE TRADITIONAL CONTEXTS OF FOLK-RNN MODELS

Curating and editing generated transcriptions—as did Banarsë in creating Let’s Have Another Gan Ainm and other musicians we variously engaged—is one mode of using our models; generating transcriptions interactively with the system is a different mode. In his piece Bastard Tunes, Ben-Tal [38] used the different generation parameters to pull the model away from the traditional conventions and used the results as precompositional material.

The parameters available for controlling generation sometimes have no direct or easily predictable effects. Setting the mode and meter can have obvious outcomes but also less obvious ones given fewer examples for the model to learn from, e.g. 9/8 in mixolydian. The folk-rnn models are also highly nonlinear; interaction between the different initialization parameters is opaque but sometimes creatively fruitful. The temperature parameter has an obvious effect: Very low temperatures (such as 0.1) will usually yield repetitive sequences. High temperatures (2.0 and above) can have dramatic effects and produce what seem like parodies of “new music” (see Fig. 2). Increasing the temperature essentially makes more symbols likely and decreases the dependency of the output at each step on what precedes it. Changing the seed allows for regenerating from the same initial settings, which is also a useful compositional tool.

Thus it is not straightforward for the composer to steer the generative process to produce “useful” outputs. While Ben-Tal’s initial interaction with the system was mostly trial and error, after generating many hundreds of outputs (discarding the vast majority), he felt able to steer the process in directions that he found compositionally useful. This turned out to be mostly through initializing the generation process with combinations uncommon in the original data, including less-common meters and modes, nonmodal opening sequences or long notes or rests (rare in these dance-based tunes). His precomposition process became an interactive search for regions of the model’s “creative space,” where the stylistic conventions modeled through the data are sufficiently weak but not entirely erased.

As in any creative work, what is useful is determined personally rather than by rule. Cocreating with folk-rnn is an act of imagination as well as iterative generation. This push and pull between composer and system can lead to new discoveries for the composer. For instance, in bars 143–145 of the first movement of Bastard Tunes (Fig. 4), the higher temperature settings led the model to produce a “jazzy” moment. Ben-Tal’s ensuing composition process involved identifying this material and choosing to bring it out in the piece. Another composer might have found it out of place and deleted or obscured it instead. The idea of composing with external constraints is, of course, not new or ground-breaking. But, as these bars illustrate, the constraints imposed by the system are not arbitrary but grounded in music. While the AI system only captures a limited aspect of musical practice, it still learns something meaningful from traces of human musical activity.
Over the summer of 2018 we organized a composition competition centered on folkrnn.org. Submissions included both a score for a set ensemble (flute, clarinet, violin, cello, piano) and an accompanying text describing how folkrnn.org contributed to the work's composition. The judging panel—Ben-Tal, Elaine Chew and Sageev Oore—considered the musical quality of each submission as well as the creative use of the model. The winning piece, Gwyl Werin by Derri Lewis, was performed by the New Music Players at a concert organized in partnership with the 2018 O'Reilly AI Conference in London. Lewis said he did not want to be "too picky" but rather selected a tune to work from after only a few iterations with folkrnn.org. He did not use the generated tune as a melodic line directly in the piece; rather he treated the generated tune as a tone row and composed harmonic, melodic and motivic material out of it.

Both Ben-Tal and Lewis used folk-rnn in a manner consistent with what Lubart described as "computer as pen pal" [39]. The process is still one sided: The computer generates ideas and the composer chooses, modifies or asks for new ideas. One possible improvement of music AI tools would be to make the process interactive, such that the computer can evaluate the individual choices and adapt what is offered in response.

CONCLUSIONS

AI will continue to be applied to and impact the domain of music. Our work demonstrates that there is an audience willing to engage with music AI. Both professional and amateur musicians have found ways of including the models we developed in their musical activities. The Web interface we deploy at folkrnn.org is a friendlier user interface than running the computer code directly. However, as we learn more about what different users find useful (and not), we aim to improve the usability of our system. We see evidence of users interactively searching for creative ideas with the system, through the iterative process of generating and altering parameters. As Lomas [40] observed, the aims of such creative search include exploring the conceptual space, identifying fruitful locations, refining ideas and seeking novelty. Translating his methods from the visual to the audio domain and from numeric to symbolic data is not straightforward. However, one of our future aims is to develop an "artificial critic." We envision not an aesthetic evaluator but rather an assistant that could facilitate exploration of the algorithm's creative space. Using similarity ratings (not a trivial question for music [41]) could help a user map out different regions of the space and home in on those more relevant (to them). Conversely, dissimilarity can be leveraged when a user decides to look for novelty or contrast. The assistant could also filter out completely unacceptable outputs based on user input for the immediate task, e.g. by building a "stylistic conformity" sieve that allows a more nuanced control of the model's adherence to training data conventions. Of course, individual users may prefer outputs that conform or those that deviate from the style of the training set.

Broadly, it is imperative that creative AI researchers engage more thoroughly with a variety of practitioners. AI has the potential to augment human creativity, and we believe such cocreativity is more fruitful than a focus on replicating (and thereafter replacing) human creativity. Such an approach is more fruitful not only for artistic creation but also for AI researchers. For instance, creative interrogation of our system [42] reveals that the "intelligence" of our AI is rather shallow and brittle. Making a technology accessible to a wider audience can also reveal new avenues for development. At the same time, demonstrating the cocreative potential of AI will also help allay some fears of this new technology.

Acknowledgments

Funding was provided as AHRC UK project no. AH/R004706/1, “Engaging three user communities with applications and outcomes of computational music creativity.”

References and Notes

11 Agres et al. [3]; Fernández and Vico [9].
15 Mossmyr et al. [12].
16 For a more detailed discussion see Sturm et al. (2016) [12].
17 Sturm and Ben-Tal [12]; Sturm et al. (2018) [12].
18 Section 3.2 in Sturm and Ben-Tal [12].
19 Section 3.3 in Sturm and Ben-Tal [12].
20 Section 3.4 in Sturm and Ben-Tal [12].
21 Implementation details can be found at www.github.com/tobyspark/folk-rnn-webapp (accessed 27 May 2021).
22 See Ref. [13].
24 Section 3.3 of Sturm and Ben-Tal [12].

32 Sturm and Ben-Tal [31].
33 Daren Banarse, email communication with Bob Sturm, 6 November 2018.
34 Ethics approval for this deception was granted by Kingston University research ethics panel. See Sturm and Ben-Tal [31].
37 Kevin McDermott, email communication with Bob Sturm, 9 September 2018.
38 See Sturm et al. (2018) [12].

ODED BEN-TAL is a composer and senior lecturer in music at Kingston University (U.K.). In addition to working with AI, he regularly uses his own intelligence to compose music.

MATTHEW TOBIAS HARRIS researches audiences and interaction, prompted by his art practice. For this project, he was a postdoctoral researcher at Queen Mary University of London, where he also helped teach a robot to perform standup comedy, and taught design to computer science and psychology students.

BOB L.T. STURM is an associate professor and PI of the MUSAiC ERC project (grant agreement No. 864189). He has been an enthusiast of Irish traditional music since living in Limerick, Ireland, during the summer of 2000.