

Co-creation and ownership for AI radio

Skylar Gordon¹, Robert Mahari^{1,2}, Manaswi Mishra¹, Ziv Epstein¹

{sfgordon, rmahari, manaswim, zive}@mit.edu

¹ MIT Media Lab ² Harvard Law School

Abstract

Recent breakthroughs in AI-generated music open the door for new forms for co-creation and co-creativity. We present Artificial.fm, a proof-of-concept casual creator that blends AI-music generation, subjective ratings, and personalized recommendation for the creation and curation of AI-generated music. Listeners can rate emergent songs to steer the evolution of future music. They can also personalize their preferences to better navigate the possibility space. As a “slow creator” with many human stakeholders, Artificial.fm is an example of how casual creators can leverage human curation at scale to collectively navigate a possibility space. It also provides a case study to reflect on how ownership should be considered in these contexts. We report on the design and development of Artificial.fm, and provide a legal analysis on the ownership of artifacts generated on the platform.

Introduction

“As notions about the nature and function of music become embedded in the structure of software-based musical systems and compositions, interactions with these systems tend to reveal characteristics of the community of thought and culture that produced them.”

—George Lewis (Lewis, 2000)

Recent breakthroughs in deep learning have introduced the opportunity for generating high-fidelity songs in the raw audio domain. Some believe that this new potential portends the end of musical creativity, while others think it represents yet another tool to augment musical production. Both sides have merit, but of particular importance is the fact that these end-to-end music generation systems can synthesize music without any understanding in music composition or technique. This opens up the possibility of embedding them in computational creativity systems, which allows users to explore a large possibility space of music without formal musical training (Compton and Mateas, 2015) and engage in *mixed-initiative co-creativity* Yannakakis, Liapis, and Alexopoulos (2014).

To interrogate these questions, we introduce a proof-of-concept casual creator system, Artificial.fm, that allows listeners to help curate and steer the evolution of music generated with OpenAI’s Jukebox model (Dhariwal et al., 2020). In addition to listening to this new kind of music, listeners

can also provide feedback on the generated songs, thus helping the AI learn to generate better music in the future. The system also uses these ratings to provide personalized music recommendations, which helps the music evolve to fit the preferences of the listener.

These components form an interconnected sociotechnical system for music generation and curation, with many distinct human stakeholders. This distributed model of production complicates the definition of *the user* of the system, since many different users are involved in different ways. It also raises important questions about who owns the artifacts generated by the system.

Our system falls in the lineage of “slow creators” defined by a “problematic gulf of execution” (Compton, 2019). This collection of creators involves most audio-based generators, since evaluating songs requires the user to actually listen to the outputs, instead of quickly discerning its quality, as with visuals. However, a distinct yet understudied aspect of Artificial.fm is the fact that generation itself is high-latency and therefore impossible to do on the fly: Jukebox takes about 20 hours to generate 20 seconds of audio. Thus Artificial.fm explores design patterns for a growing set of systems where intensive underlying computation means real-time interaction with the underlying generator is fundamentally infeasible.

In this paper, we present the case study of Artificial.fm to highlight how slow creation can translate to the evaluation and curation of AI-generated music. We then use legal precedent to trace the multiple stakeholders involved in this process and unpack the each actor’s stake in ownership.

Related Work

Algorithmic Music has a rich history amongst composers starting in the pre-computing era from the process works of George Brecht’s Drip Music (1962), Stockhausen’s Setz die Segel zur Sinn (Maconie, 1970) and Xenakis’ Formalised Music (Xenakis, 1992) to the formation of the US League of Automatic Music Composers (1978).

Algorithms have been used to generate music both in the symbolic domain (Hiller Jr and Isaacson, 1957; Moorer, 1972; Hadjeres, Pachet, and Nielsen, 2017; Huang et al., 2018) and in the waveform domain through digital vocoders (Bonada and Serra, 2007; Blaauw and Bonada, 2017) and synthesizers (Mehri et al., 2016; Engel et al., 2017).

Compton (2019) identifies Musical computational creativity systems as inherently “slow creators” where the user evaluation in the *grokloop* is implicitly slow. Though computationally slow in generating these more complex musical generative spaces, this newer generation of ‘slow systems’ are capable of producing more aesthetically pleasing and uniquely shaped outputs that often feel more rewarding and personal to the user. More recent casual creators like -Magenta’s Tone Transfer, Piano Genie and applications produced at the BitRate ML and Music hackathon 2020, take advantage of modern AI models to produce higher fidelity musical outputs.

These generators are characterized by large possibility spaces, which can be difficult for individuals to explore. A promising approach to rapidly search through a large possibility space to find the “gem in the rough” is to crowdsource its exploration. A diverse set of casual creators leverage collaborative media to produce intriguing artifacts. The Reddit R/Place experiment had users collaboratively paint a pixel canvas (Rappaz et al., 2018). Drave’s Electric Sheep used user feedback and evolutionary algorithms to generate new “sheep” - fractal animations adapted to crowd preferences (Draves, 2005). PicBreeder also uses evolutionary algorithms and allows users to collaboratively evolve images Secretan et al. (2008, 2011). Feed the Ganimals allowed users to explore and curate AI-generated hybrid animals, and found that social cues led to the formation of diverse local trends (Epstein et al., 2020a, 2021).

System Overview

Artificial.fm uses OpenAI’s Jukebox (Dhariwal et al., 2020), a deep neural network trained on 1.2 million songs, for music generation. Jukebox has the ability to take as input a “prime” of existing music which it then improvises on top of. We solicit primes from local musicians as part of a collaboration to support artists affected by the pandemic. Jukebox also takes in a specified artist and genre as inputs which condition the style of the generated song outputs.

The outputs of the song generation process are streamed via the platform, where listeners provide subjective feedback on the AI-generated music, in the form of ratings. The questions related to how happy, danceable, artificial, instrumental, upbeat and song was, and how clear the lyrics, and if they liked it, on a 5-point Likert scale (see Supplementary Information Section 1.2 for more information).

The crowdsourced feedback is then used to adapt the generation process with an algorithm that balances exploring new permutations of parameters with exploiting existing parameters that are popular with users. This is achieved using a variation of Thompson sampling, which is regret-minimizing in such contexts (Chapelle and Li, 2011). To do so, we use the Spotify API and Essentia (Bogdanov and others, 2013) to generate a rich set of covariates for the artist of the prime, as well as candidate artists and genres (see Supplementary Information Section 1.3 for more information on how these covariates are generated).

As new primes are solicited from local musicians, the following algorithm finds parameters (e.g. an artist and genre prompt) to pair with that prime to balance exploration and

exploitation: First, we fit a model \hat{f} predicting ratings of the existing songs (e.g. How much do you like this song?) based on the Spotify covariates of that song’s prime artist, artist prompt, and genre prompt (27 features total). Then, we sample M artist, genre pairs from the joint distribution of these prompts in the input space. Then, we predict the rating of that artist, genre pair for the given prompt $\hat{f}_{prime}(a_\ell, g_\ell)$. We then take the top γ artist, genre pairs and randomly sample one uniformly (here γ controls exploration vs exploitation, $\gamma = 1$ is maximal exploitation, $\gamma = M$ is maximal exploration). See Supplementary Information Section 1.4 for more details about this algorithm.¹

Artificial.fm also provides personalized song recommendations to users. Through a preference elicitation interface (see Figure S3), users can explicitly specify the kind of songs they would like to hear. Based on their stated preferences, a personalized recommendation algorithm serves songs to them consistent with these preferences (see Supplementary Information Section 1.5 for more details about this recommender system).

Data and Results

As of July 21, 2021, we accumulated 522 ratings of 71 songs by 40 people. The songs were generated with genre prompts from folk, house, pop, americana, rock, classical, electronic, and funk, and artist prompts from The Weeknd, Aerosmith, The Doors, Justin Bieber, Elton John, Dolly Parton, Otis Redding, and Lady Gaga. The primes were sourced from several local artists we reached out to. The 40 people found the platform through word of mouth.

The distribution of ratings by question is shown on the left of Figure 1. Relative to the other questions, listeners found the songs highly artificial (one-sided t test, $p < 0.001$), and lacking in clear lyrics ($p < 0.001$). This suggests that the music of Artificial.fm may not fall into the “normal distribution” of what you find on the radio, but instead represents a polyphonous new kind of music onto itself. That being said, we did find meaningful variation in all seven questions ratings, which suggests there is quantifiable diversity in the possibility space to explore and optimize.

The pairwise correlations between these seven questions are shown on the right of Figure 1. We observe that perceptions of liking a song is associated with ratings of a song being danceable ($R = 0.75$, $p < 0.001$), instrumental ($R = 0.44$, $p = 0.004$), and having clear lyrics ($R = 0.35$, $p = 0.037$). We also find that ratings of the artificiality of a song are marginally negatively associated with having clear lyrics ($R = -0.30$, $p = 0.088$), and how happy the song is perceived to be ($R = -0.29$, $p = 0.078$).

¹This algorithm assumes there is already a large number of both songs and ratings and therefore requires solving the “cold-start problem.” Since the scope of this short paper is introducing the concept of AI radio via Artificial.fm with preliminary user testing and ethical considerations, this algorithm should be considered as a sketch for how Artificial.fm would work at scale. As such, we leave formal evaluation of such an approach to music generation to future work.

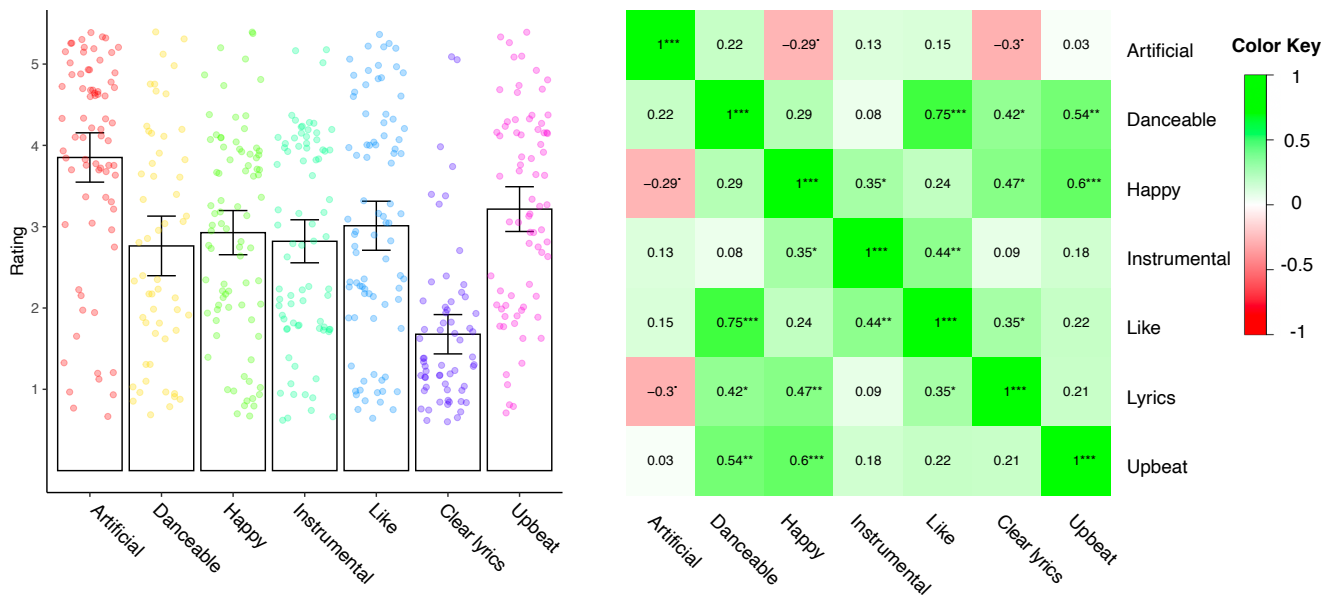


Figure 1: Left: Song ratings by question. Right: Pairwise correlation matrix between question ratings. · refers to $p \leq 0.1$, * refers to $p \leq .05$, ** refers to $p \leq .01$, *** refers to $p \leq .001$.

Ownership of AI-Generated Music

The owner of a casual creators’ output should be the entity responsible for creation. Compton and Mateas (2015) defines casual creation as the “the creation [of] new artifacts that bring feelings of pride, ownership, and creativity to the users that make them.” This definition centers the users of a casual creator as the owners of its output. Artificial.fm challenges this idea of ownership and highlights open questions related to ownership of AI-generated works.²

At least five actors could claim some level of ownership over the works created by Artificial.fm: (1) the artist who submitted the prime on which a piece of music is based, (2) the many artists whose music was used to train Jukebox, (3) the system architects who developed Artificial.fm, (4) the listeners whose ratings are used to steer music production, and (5) the artificial intelligence itself. This section begins to explore the question of ownership for casual creators by analyzing the legal basis on which these actors may claim ownership and concludes by suggesting ownership models better suited to the distributed nature of systems like Artificial.fm.

The Prime and Training Artists 🧑🏻‍🎤🧑🏻‍🎤

Both the prime and training data artists could claim ownership over a given piece of music created by Artificial.fm by arguing that Artificial.fm infringes on their copyright. To this end, they would need to show that the generated music is “substantially similar” to their work (Williams v.

²We use the term “ownership” broadly to encompass all the rights commonly associated with authorship. Where relevant, this section will base its analysis on U.S. and California law.

Gaye, 2018) and that the music was not independently created (Feist v. Rural Telephone Service Co., 1991).

The prime artists explicitly provide direct access to their works but the training data artists do not. Moreover, it is unclear whether Artificial.fm has “access” to the underlying training data because the music in the training data has been transformed into the Jukebox algorithm which does not contain copies of the works it has been trained on. Even if Artificial.fm has access to the artists’ work, an infringement claim would require showing that a song created by Artificial.fm is substantially similar to a given artist’s work (Williams v. Gaye, 2018). Substantial similarity is assessed using a two part test: first, an objective test where a music expert analytically compares the elements of two works for substantial similarity and second, a subjective test where an “ordinary reasonable person” assesses if the two works feel substantially similar (Swirsky v. Carey, 2004). Different experts and “ordinary people” may disagree about substantial similarity making these tests inherently vague. In the Artificial.fm case, it is likely that some generated music is similar to some works owned by prime artists, but it is unlikely for generated music to be substantially similar to songs in the training data.

The System Architects and Listeners 🧑🏻‍🔧🧑🏻‍🔧

The system architects and listeners play their own role in creating the output of Artificial.fm and could claim ownership over the generated content. To focus on their contribution, imagine that Artificial.fm was trained exclusively on works in the public domain.

On one hand, the system architects might be akin to photographers who compose photographs by documenting objects from the real world. The U.S. Supreme Court clari-

fied in 1884 that photography is to be treated as an art under copyright law, and that the photographer is to be treated as the “mastermind” whose creativity gives rise to a copyrightable work (Burrow-Giles Lithographic Co. v. Sarony, 1884). The system architects can similarly be characterized as the masterminds, who use their ingenuity to take advantage of a technology to produce works of art. On the other hand, although the listeners are using a tool built by the system architects, it is the listeners’ preferences, not the architects’, that guide what Artificial.fm produces. In this sense, the listeners are akin to photographers and the system architects are similar to camera makers, who have no claim to the photographs made with the technology they built.

Along these lines, the AI Artist Mario Klingemann often refers to himself as a “neurographer,” a photographer of neural landscapes (Castelle, 2020). Artificial.fm employs several design patterns so that listeners can earnestly explore the possibility space, and hence become neurographers of sorts. The personalized song recommender and preference pane push the onus of creativity onto the listener, which may in turn strengthen their ownership claim.

The Artificial Intelligence Itself 🧠

Perhaps the true author of Artificial.fm music is the AI (United States Copyright Office, 2021). Like a photographer, the AI decides what to create based on underlying criteria and thus identifies a small subset of expressions from a large pool of possibilities. In support of this idea, Colton et al. (2020) present the framework of the *machine condition*, by which machines creatively express their own subjectivity. However, the AI could also be compared to a sophisticated camera, a tool to enable others to create art without contributing creativity itself. Tracing the history of photography and animation, Hertzmann (2018) advances this idea and argues that only *social agents* can create art. Epstein et al. (2020b) find that there is natural heterogeneity in the extent to which people anthropomorphize AI (i.e. think of it as a tool vs an agent), and that these perceptions of agency are related to allocations of responsibility and credit for the involved human stakeholders.

If the AI *is* capable of creativity, this raises the question of whether it is “working” for whoever built it or whether it is autonomous. In the former case, the original creator of the AI might own any creative expression created by it (under the work for hire principle (Bridy, 2012)). In the latter case, the AI might exist as some form of DAO (decentralized autonomous organization) that could be capable of ownership.

A Distributed Approach to Ownership

Likely for pragmatic reasons, traditional copyright law favors resolutions with a small number of copyright owners. Many actors contribute to Artificial.fm in distinct ways, and so traditional ownership norms may be an ill fit. As a result, Artificial.fm, and platforms like it, do not fit neatly into existing ownership norms and are more suited to a distributed ownership model that divides ownership among all the actors involved in the process of casual creativity. Data cooperatives and non-fungible tokens (NFTs) are two possible

technical approaches to such ownership structures.

A data cooperative is a member-owned entity, similar to a credit union, that administers data voluntarily pooled by its members to safeguard data rights, protect privacy, and facilitate data monetization (Pentland and Hardjono, 2020). While data cooperatives are usually associated with personal data, they may also be useful in the context of casual creators, where all the actors who contribute to the creation of a set of works pool these works in a cooperative that advocates on behalf of all the creators.

NFTs are an application of blockchain ledgers to track the ownership of unique digital assets, which facilitates a large number of owners. In the casual creators context, all actors involved in the creative process could receive NFTs that give them fractional ownership over one or more works.

Both data cooperatives and NFTs are technical solutions to facilitate distributed ownership, but neither solution provides an answer to how much ownership each actor *should* receive. The normative question of how to allocate this ownership fairly and in a way that incentivizes casual creativity, is beyond the scope of this paper, but remains an open and exciting question for our community.

Conclusion

In leveraging AI for song generation, one might wonder if a formula for good music emerges. In using users’ preference for songs as a metric for how good songs are, what music is perceived to be better is considerably unpredictable, making it difficult to optimize AI systems to generate “good” music that people enjoy listening to. Indeed, much of the time music’s perceived quality is closely related with its popularity (Salganik, Dodds, and Watts, 2006). With music’s social context being extremely influential to the public’s opinion of what is good music, and gives rise to a snowball effect of “the rich get richer,” as the more popular songs gain more popularity while less popular songs do not see the same increase in streaming. As such, the design of the system becomes increasingly important, both to calibrate the listener’s expectations for the music they will hear, and to surface cues necessary for them to make informed decisions. Casual creators like Artificial.fm bring us one step closer to understanding and integrating social context into AI systems, which in turn bootstraps their creative potential.

Acknowledgements

We would like thank Max Kreminski, Amy Smith, Simon Colton, Sandy Pentland, Abhimanyu Dubey, Christopher Bavitz, Jessica Fjeld, Daza Greenwood, Matt Groh and Jack Muller for helpful feedback and discussion. We would also like to thank the musicians who provided primes to Artificial.fm, and the listeners of Artificial.fm, as well as the artists whose music Jukebox was trained on.

Author Contributions

SG, MH, MM and ZE conceptualized the project. SG and ZE developed the system. SG, MH, MM and ZE wrote the paper.

References

- Burrow-Giles Lithographic Co. v. Sarony. 1884. Supreme court of the united states.
- Blaauw, M., and Bonada, J. 2017. A neural parametric singing synthesizer. *arXiv preprint arXiv:1704.03809*.
- Bogdanov, D., et al. 2013. Essentia: An audio analysis library for music information retrieval. In *14th Conference of the International Society for Music Information Retrieval*. ISMIR.
- Bonada, J., and Serra, X. 2007. Synthesis of the singing voice by performance sampling and spectral models. *IEEE signal processing magazine* 24(2):67–79.
- Bridy, A. 2012. Coding creativity: copyright and the artificially intelligent author. *Stan. Tech. L. Rev.* 5.
- Castelle, M. 2020. The social lives of generative adversarial networks. In *FAT**, 413.
- Chapelle, O., and Li, L. 2011. An empirical evaluation of thompson sampling. *Advances in neural information processing systems* 24:2249–2257.
- Colton, S.; Pease, A.; Guckelsberger, C.; McCormack, J.; Llano, T.; et al. 2020. On the machine condition and its creative expression. In *International Conference on Computational Creativity*.
- Compton, K., and Mateas, M. 2015. Casual creators. In *ICCC*, 228–235.
- Compton, K. 2019. *Casual creators: Defining a genre of autotelic creativity support systems*. University of California, Santa Cruz.
- Dhariwal, P.; Jun, H.; Payne, C.; Kim, J. W.; Radford, A.; and Sutskever, I. 2020. Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341*.
- Draves, S. 2005. The electric sheep screen-saver: A case study in aesthetic evolution. In *Workshops on Applications of Evolutionary Computation*, 458–467. Springer.
- Engel, J.; Resnick, C.; Roberts, A.; Dieleman, S.; Norouzi, M.; Eck, D.; and Simonyan, K. 2017. Neural audio synthesis of musical notes with wavenet autoencoders. In *International Conference on Machine Learning*, 1068–1077. PMLR.
- Epstein, Z.; Boulais, O.; Gordon, S.; and Groh, M. 2020a. Interpolating gans to scaffold autotelic creativity. *arXiv preprint arXiv:2007.11119*.
- Epstein, Z.; Levine, S.; Rand, D. G.; and Rahwan, I. 2020b. Who gets credit for ai-generated art? *Isience* 23(9):101515.
- Epstein, Z.; Groh, M.; Dubey, A.; and Pentland, A. 2021. Social influence leads to the formation of diverse local trends. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW2):1–18.
- Feist v. Rural Telephone Service Co. 1991. Supreme court of the united states.
- Hadjeres, G.; Pachet, F.; and Nielsen, F. 2017. Deepbach: a steerable model for bach chorales generation. In *ICML*, 1362–1371. PMLR.
- Hertzmann, A. 2018. Can computers create art? In *Arts*, volume 7, 18. Multidisciplinary Digital Publishing Institute.
- Hiller Jr, L. A., and Isaacson, L. M. 1957. Musical composition with a high speed digital computer. In *Audio Engineering Society Convention 9*. Audio Engineering Society.
- Huang, C.-Z. A.; Vaswani, A.; Uszkoreit, J.; Shazeer, N.; Simon, I.; Hawthorne, C.; Dai, A. M.; Hoffman, M. D.; Dinculescu, M.; and Eck, D. 2018. Music transformer. *arXiv preprint arXiv:1809.04281*.
- Lewis, G. E. 2000. Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal* 33–39.
- Maconie, R. 1970. Stockhausen’s ‘setz die segel zur sonne’. *Tempo* (92):30–32.
- Mehri, S.; Kumar, K.; Gulrajani, I.; Kumar, R.; Jain, S.; Sotelo, J.; Courville, A.; and Bengio, Y. 2016. SampleRNN: An unconditional end-to-end neural audio generation model. *arXiv preprint arXiv:1612.07837*.
- Moorer, J. A. 1972. Music and computer composition. *Communications of the ACM* 15(2):104–113.
- Pentland, A., and Hardjono, T. 2020. 2. data cooperatives. In *Building the New Economy*. 0 edition. <https://wip.mitpress.mit.edu/pub/pnxgvubq>.
- Rappaz, J.; Catasta, M.; West, R.; and Aberer, K. 2018. Latent structure in collaboration: the case of reddit r/place. In *Twelfth International AAAI Conference on Web and Social Media*.
- Salganik, M. J.; Dodds, P. S.; and Watts, D. J. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.
- Secretan, J.; Beato, N.; D’Ambrosio, D. B.; Rodriguez, A.; Campbell, A.; and Stanley, K. O. 2008. Picbreeder: evolving pictures collaboratively online. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1759–1768.
- Secretan, J.; Beato, N.; D’Ambrosio, D. B.; Rodriguez, A.; Campbell, A.; Folsom-Kovarik, J. T.; and Stanley, K. O. 2011. Picbreeder: A case study in collaborative evolutionary exploration of design space. *Evolutionary computation* 19(3):373–403.
- Swirsky v. Carey. 2004. United states court of appeals for the ninth circuit.
- United States Copyright Office. 2021. Compendium of us copyright office practices.
- Williams v. Gaye. 2018. United states court of appeals for the ninth circuit.
- Xenakis, I. 1992. *Formalized music: thought and mathematics in composition*. Pendragon Press.
- Yannakakis, G. N.; Liapis, A.; and Alexopoulos, C. 2014. Mixed-initiative co-creativity.