

# DeepDrake ft. BTS-GAN and TayloRVC: An Exploratory Analysis of Musical Deepfakes and Hosting Platforms

Michael Feffer, Zachary C. Lipton and Chris Donahue

Carnegie Mellon University, Pittsburgh, PA, US

## Abstract

Recent advancements in voice conversion and text-to-speech technology have facilitated the creation of *musical deepfakes*, audio tracks featuring the voices of celebrity artists—typically without the artists’ involvement. Several deepfakes have already gone viral, leaving the music industry scrambling to sort out the potential impacts. While the media have primarily focused on specific high-profile incidents, there has been less attention from journalists and researchers surrounding the broader trends in musical deepfakes, including the communities creating them, the modeling techniques that they employ, and the sites on which they congregate. In this paper, we investigate two leading sources of musical deepfake models, the AI Hub Discord server and the Uberduck website, which are dedicated to the training, utilization, and distribution of these deepfakes. Interestingly, musical deepfakes target hundreds of artists of different backgrounds, levels of success, and musical styles. In light of the economic, legal, and ethical issues raised by deepfakes of so many artists, we provide warnings about the generation of discriminatory forms of content and potential financial and contractual problems for artists. We recommend more research should be conducted in this area, especially to probe peoples’ perceptions of this technology and devise approaches that mitigate potential harms.

## Keywords

Deepfake, GAN synthesis, Diffusion models, Artist identity and representation

Recent progress in speech and music processing research has yielded tools that can realistically imitate the voices of famous music artists. As a result, people with only modest technical and musical skills can now create *musical deepfakes*, songs featuring the likeness of an artist, namely their voice, often without the knowledge or consent of the artist. Though the existence of image and video deepfakes is now common knowledge [1, 2, 3], musical deepfakes are a novel phenomenon. Artists stand to lose both financially and reputationally from songs that coopt their identities. Labels and streaming services may find themselves in competition with AI forms of artists from monetary and legal standpoints. Moreover, the ability to copy musical styles and impersonate any artist, living or dead, poses questions of ethics and identity [4, 5, 6].

A driving force behind the sudden escalation of musical deepfaking is the ease with which they can now be created, controlled, and shared. Generative modeling of broad music audio (i.e., complete songs) remains challenging—a key simplification behind the present proliferation

---


HCMIR23: 2nd Workshop on Human-Centric Music Information Research, November 10th, 2023, Milan, Italy

✉ mfeffer@andrew.cmu.edu (M. Feffer); zlipton@andrew.cmu.edu (Z. C. Lipton)

🌐 <https://mfeffer.github.io> (M. Feffer); <https://www.zacharylipton.com> (Z. C. Lipton); <https://chrisdonahue.com> (C. Donahue)

🆔 0000-0002-5243-472X (M. Feffer); 0000-0002-3824-4241 (Z. C. Lipton); 0009-0007-6825-6327 (C. Donahue)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

of musical deepfakes is modeling the more narrow distribution of individual *singing voices*. Compared to modeling broad music audio [7, 8, 9, 10] which requires new ML methods, thousands of hours of training data, and specialized hardware, modeling singing voice is possible with off-the-shelf methods, minutes of training data, and commodity hardware. To turn generated vocals into a complete song, model outputs are combined with manually-created musical elements (e.g., mixing deepfaked rap with a human-composed beat).

There are two broad categories of approaches for singing voice synthesis (SVS): (1) *voice conversion* (VC), and (2) *text-to-speech* (TTS), primarily differentiated by the forms of user control they offer—respectively, VC is controlled by singing and TTS is controlled by lyrics represented as text. Both categories involve training models which estimate singing audio from intermediary features: VC-based models use intermediaries that can be readily extracted from input singing such as fundamental frequency [11, 12] or representations from pre-trained encoders [13], while TTS-based models use lyrics (and sometimes melody notes). In both cases, intermediary features both simplify the modeling problem (thereby decreasing compute and data requirements) and afford an essential form of control for musical deepfakes: the ability to specify lyrics (either by singing or writing text). Popular SVS systems are complex pipelines which compose several modules for feature extraction [14, 15, 13] and resynthesis [16, 17, 18]. Despite the underlying complexity, training and using models is made more broadly accessible by the distribution of easy-to-use open source tools<sup>1</sup> and video tutorials.<sup>2</sup>

Even with such accessible resources, technical and musical expertise are still required to train and co-create with singing voice models. Hence, making convincing musical deepfakes is, for the moment, primarily accessible to musical “prosumers” (e.g., “bedroom producers” already familiar with technical music production tools). Additionally, considerable artistic effort—composing and performing lyrics and producing backing tracks—is also a requirement. Despite these impediments, music streaming services are already being flooded with musical deepfakes [19, 20, 21]. Such deepfakes have also gone viral on social media [22], prompting everyone from listeners to musicians and record labels to seriously consider the issues these capabilities raise [23]. Moreover, musical deepfakes may become even easier to create in the future—the recent and rapid advancement in broad music audio generation methods suggest that it may eventually be possible for anyone to generate convincing musical deepfakes without technical or musical expertise.

Analysis of initial trends in musical deepfaking, such as examining which types of artists have been targeted, can help navigate these dilemmas or better prepare for future developments. Surprisingly, except for examples that have gone viral, we find little coverage in that regard. To this end, we explored AI Hub, a Discord community at the center of musical deepfake creation [24], and scraped the website Uberduck.ai<sup>3</sup> (referred to as “Uberduck” going forward) in order to gather information on current deepfake models. While AI Hub is a community effort driven by prosumers sharing models, Uberduck is backed by a corporation and hosts models that require comparatively less technical background. Our results suggest that hundreds of musicians of diverse backgrounds have stakes in these issues. Based on our analysis, we also

---

<sup>1</sup>SoVITS and RVC are popular tools for VC: <https://github.com/voicepaw/so-vits-svc-fork>, <https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI/tree/main>.

<sup>2</sup>Example video tutorial: <https://www.youtube.com/watch?v=tZn0lcGO5OQ>

<sup>3</sup><https://uberduck.ai/>

offer recommendations for the future, including but not limited to research into perceptions of listeners and members of the music industry as per Lee et al. [25].

## 2. Methodology

We gathered data from AI Hub by recording the title, post date, and tags of all posts in the voice-models channel on May 31st, 2023. This means we gathered all posts ever made in the channel from the Discord’s inception to May 31st. Regarding Uberduck, we similarly downloaded details of all available voice models on the site as of May 31st in the form of JSON metadata. For each data source, we first manually labeled entries with relevant artist info, including the artist’s name, race<sup>4</sup>, and whether the artist is deceased. We then used APIs for MusicBrainz [26] and Spotify to gather additional data about each artist’s gender<sup>5</sup>, music genres, geographical region, and popularity on a scale from 0 to 100 (with 100 being most popular).<sup>6</sup>

## 3. Analysis

Overall, we found that nearly 400 artists were represented in AI Hub models, and over 50 were represented in Uberduck models. Additionally, for the first four weeks of May 2023, over 100 model posts were made per week in AI Hub. Based on retrieved metadata, users made different models to utilize differing training approaches (e.g., SoVITS versus RVC) or capture artists at different points in their careers (e.g., early versus contemporary Britney Spears). Table 1a displays the ten most popular artists from AI Hub and Uberduck in terms of how many models were made using their data, and Table 1b displays results of a random sample of models from each source. Evidently, the most popular artists in AI Hub typically have more related models than those in Uberduck. However, both lists highlight artists from a wide range of musical styles and backgrounds. In particular, Juice WLRD appears next to Jungkook of BTS in one list, and Kanye West and David Bowie appear in another.

The diversity of both data sources are also quantitatively illustrated in Figures 1 and 2. Namely, Figure 1 shows the distribution of artists with deepfake models in each source grouped by race, and Figure 2 does the same in each source grouped by gender, popularity score, and region. We find that AI Hub has greater diversity across each criterion, featuring a bimodal popularity score distribution and many artists from Europe and Asia. However, very popular Black American male artists are most represented in each data source. The dominance of rap and hip hop styles in Figure 3 showing the top 10 most popular artists’ genres in each source also supports this.

## 4. Discussion

While our work sheds light on the deepfake models that currently exist, we emphasize that it is, at best, a preliminary investigation. For instance, we only focus on the number of models pertaining

---

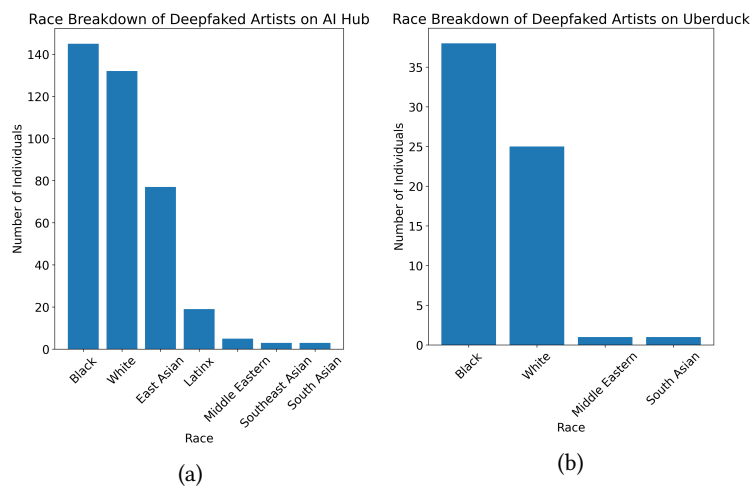
<sup>4</sup>Determined via sources ranging from physical appearance to heritage. We are aware of limitation that race is a social construct. Our aim is to illustrate diversity of impacted artists.

<sup>5</sup>This was largely provided by MusicBrainz but was occasionally inferred from photos and articles in a manner similar to that employed for race. As such, we also emphasize awareness of gender as a social construct and again stress that our aim is to showcase the range of those affected.

<sup>6</sup>Resulting data available here: [https://docs.google.com/spreadsheets/d/1tZa9YsTiFIYCF-gIndquFFnMNV\\_50TD81EFf4Z95ajE/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1tZa9YsTiFIYCF-gIndquFFnMNV_50TD81EFf4Z95ajE/edit?usp=sharing)

Artist (no. of models)		Artist from Random Sample	
AI Hub	Uberduck	AI Hub	Uberduck
Michael Jackson (10)	Eminem (5)	Duki	Damon Albarn
Juice WRLD (7)	Playboi Carti (3)	Noa Kirel	Noel Gallagher
Playboi Carti (6)	Juice WRLD (3)	Trent Reznor	Nicki Minaj
Eminem (5)	B La B (2)	Weird Al	Kanye West
Notti Osama (5)	E-40 (2)	Kendrick Lamar	NLE Choppa
Ariana Grande (4)	Freddie Mercury (2)	Winter	Lil Uzi Vert
Britney Spears (4)	Lady Gaga (2)	Killy	Liam Gallagher
Irene (4)	Lil Uzi Vert (2)	Jhene Aiko	Andy Bell
Jungkook (4)	XXXTentacion (2)	Ice Spice	MC Ride
Kanye West (4)	21 Savage (1)	Lil Tjay	David Bowie

**Table 1** Lists of artists that have deepfake models in AI Hub and Uberduck. (a) shows the top ten artists in terms of number of imitating models from each source. (b) shows ten artists based on a random sample of models from each source. Both types of lists showcase the range of represented artists in each source.

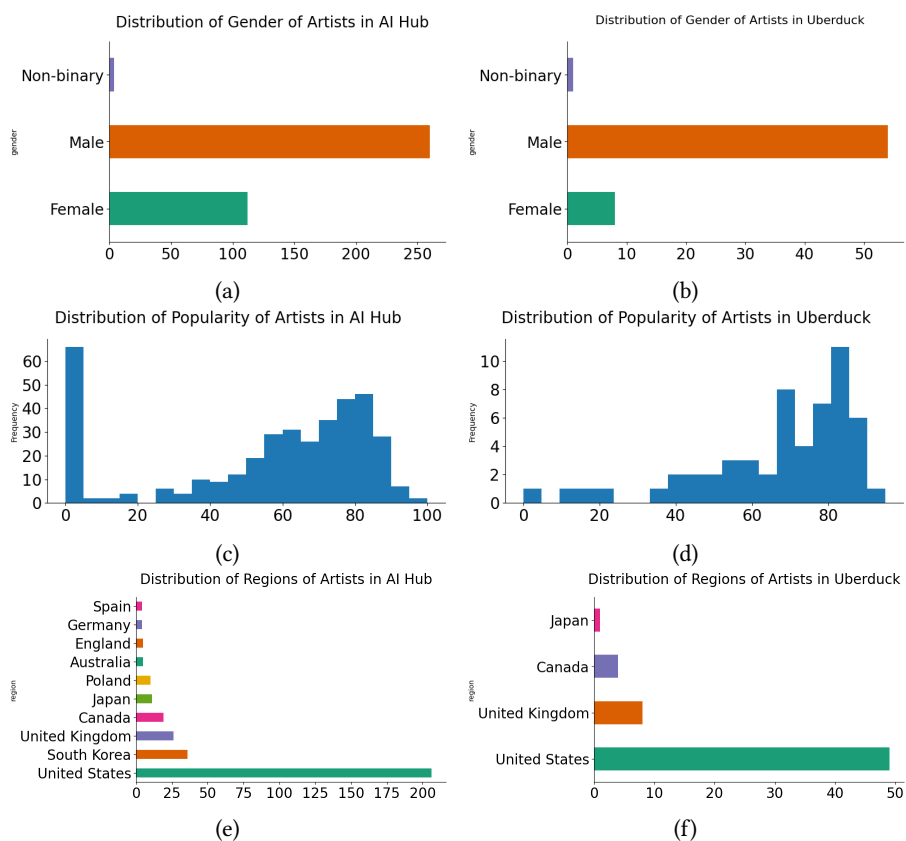


**Figure 1:** Race distribution of artists with deepfake models from each data source. (a) illustrates the AI Hub distribution while (b) illustrates the Uberduck distribution. Black artists are most represented.

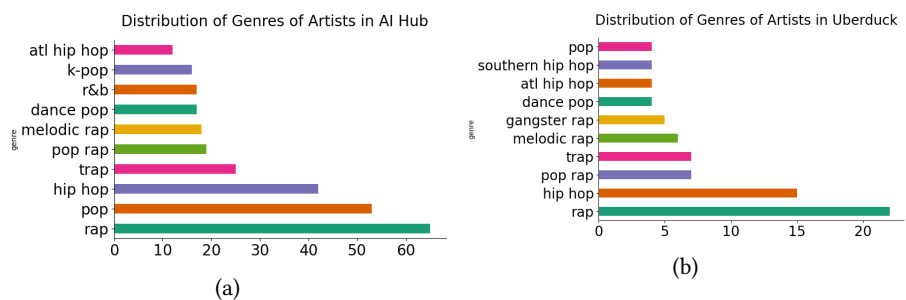
to each artist, but the numbers of songs generated would also be valuable information.<sup>7</sup>

Even so, our findings suggest some concerning possibilities. First, the usage of models imitating East Asian and Black artists by creators who do not share those demographics could be considered digital forms of yellowface and blackface respectively [27, 5]. Similarly, voice and text-

<sup>7</sup>We briefly studied other parts of AI Hub and observed originals and covers channels where users shared original tracks or covers of existing songs made with deepfakes, respectively, and each channel appeared to have hundreds of messages exchanged in a given day. Further analysis of these channels could address some of these limitations.



**Figure 2:** Additional information of artists in each source. (a), (c), and (e) correspond to gender, popularity, and region of artists in AI Hub (regions limited to top 10), and (b), (d), and (f), correspond to same criteria in Uberduck.



**Figure 3:** Genres of artists in each source. (a) corresponds to AI Hub and (d) corresponds to Uberduck.

to-speech models imitating deceased artists broach ethical and normative questions regarding whether impersonation of the dead is appropriate. Deepfake models may also exacerbate issues of music ownership as musicians already have a tenuous grasp on music ownership (see, e.g., [28, 29]). Therefore, we recommend that more research should be done in this area.

## References

- [1] L. Verdoliva, Media forensics and deepfakes: An overview, *IEEE Journal of Selected Topics in Signal Processing* 14 (2020) 910–932. doi:10.1109/JSTSP.2020.3002101.
- [2] M. Albahar, J. Almalki, *Journal of Theoretical and Applied Information Technology* 97 (2019) 3242–3250.
- [3] Y. Mirsky, W. Lee, The creation and detection of deepfakes: A survey, *ACM Computing Surveys* 54 (2022) 1–41. doi:10.1145/3425780.
- [4] M. Tracy, A ‘virtual rapper’ was fired. questions about art and tech remain., *The New York Times* (2022). URL: <https://www.nytimes.com/2022/09/06/arts/music/fn-meka-virtual-ai-rap.html>.
- [5] F. Sobande, Spectacularized and branded digital (re)presentations of black people and blackness, *Television & New Media* 22 (2021) 131–146. doi:10.1177/1527476420983745.
- [6] I. Bonifacic, “lost tapes of the 27 club” used google ai to “write” a new nirvana song, 2021. URL: <https://www.engadget.com/over-the-bridge-lost-tapes-of-the-27-club-223000315.html>.
- [7] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, I. Sutskever, Jukebox: A generative model for music, arXiv:2005.00341 (2020).
- [8] A. Agostinelli, T. I. Denk, Z. Borsos, J. Engel, M. Verzetti, A. Caillon, Q. Huang, A. Jansen, A. Roberts, M. Tagliasacchi, et al., MusicLM: Generating music from text, arXiv:2301.11325 (2023).
- [9] C. Donahue, A. Caillon, A. Roberts, E. Manilow, P. Esling, A. Agostinelli, M. Verzetti, I. Simon, O. Pietquin, N. Zeghidour, et al., SingSong: Generating musical accompaniments from singing, arXiv:2301.12662 (2023).
- [10] J. Copet, F. Kreuk, I. Gat, T. Remez, D. Kant, G. Synnaeve, Y. Adi, A. Défossez, Simple and controllable music generation, arXiv:2306.05284 (2023).
- [11] M. Morise, H. Kawahara, H. Katayose, Fast and reliable f0 estimation method based on the period extraction of vocal fold vibration of singing voice and speech, in: *Audio Engineering Society Conference: 35th International Conference: Audio for Games, 2009*. URL: <http://www.aes.org/e-lib/browse.cfm?elib=15165>.
- [12] J. W. Kim, J. Salamon, P. Li, J. P. Bello, Crepe: A convolutional representation for pitch estimation, in: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2018, pp. 161–165.
- [13] W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, A. Mohamed, Hubert: Self-supervised speech representation learning by masked prediction of hidden units, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021) 3451–3460. doi:10.1109/TASLP.2021.3122291.
- [14] B. van Niekerk, M.-A. Carbonneau, J. Zaïdi, M. Baas, H. Seuté, H. Kamper, A comparison of discrete and soft speech units for improved voice conversion, in: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, p. 6562–6566. doi:10.1109/ICASSP43922.2022.9746484.
- [15] K. Qian, Y. Zhang, H. Gao, J. Ni, C.-I. Lai, D. Cox, M. Hasegawa-Johnson, S. Chang, Contentvec: An improved self-supervised speech representation by disentangling speakers, in: *Proceedings of the 39th International Conference on Machine Learning*, PMLR, 2022, p.

- 18003–18017. URL: <https://proceedings.mlr.press/v162/qian22b.html>.
- [16] X. Wang, J. Yamagishi, Using cyclic noise as the source signal for neural source-filter-based speech waveform model, in: *Interspeech 2020, ISCA*, 2020, p. 1992–1996. URL: [https://www.isca-speech.org/archive/interspeech\\_2020/wang20u\\_interspeech.html](https://www.isca-speech.org/archive/interspeech_2020/wang20u_interspeech.html). doi:10.21437/Interspeech.2020-1018.
- [17] J. Kong, J. Kim, J. Bae, Hifi-gan: Generative adversarial networks for efficient and high fidelity speech synthesis, in: *Advances in Neural Information Processing Systems*, volume 33, Curran Associates, Inc., 2020, p. 17022–17033. URL: <https://proceedings.neurips.cc/paper/2020/hash/c5d736809766d46260d816d8dbc9eb44-Abstract.html>.
- [18] J. Liu, C. Li, Y. Ren, F. Chen, Z. Zhao, Diffsinger: Singing voice synthesis via shallow diffusion mechanism, *Proceedings of the AAAI Conference on Artificial Intelligence* 36 (2022) 11020–11028. doi:10.1609/aaai.v36i10.21350.
- [19] M. Savage, Deezer: Streaming service to detect and delete 'deepfake' ai songs, *BBC News* (2023). URL: <https://www.bbc.com/news/entertainment-arts-65792580>.
- [20] A. Johnson, Spotify removes 'tens of thousands' of ai-generated songs: Here's why, *Forbes* (2023). URL: <https://www.forbes.com/sites/ariannajohnson/2023/05/09/spotify-removes-tens-of-thousands-of-ai-generated-songs-heres-why/?sh=601d69624f4a>.
- [21] A. Hoover, Spotify has an ai music problem—but bots love it, *Wired* (2023). URL: <https://www.wired.com/story/spotify-ai-music-robot-listeners/>.
- [22] J. Coscarelli, An a.i. hit of fake 'drake' and 'the weeknd' rattles the music world, *The New York Times* (2023). URL: <https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html>.
- [23] E. Livni, L. Hirsch, S. Kessler, Who owns a song created by a.i.?, *The New York Times* (2023). URL: <https://www.nytimes.com/2023/04/15/business/dealbook/artificial-intelligence-copyright.html>.
- [24] C. Xiang, Inside the discord where thousands of rogue producers are making ai music, 2023. URL: <https://www.vice.com/en/article/y3wdj7/inside-the-discord-where-thousands-of-rogue-producers-are-making-ai-music>.
- [25] K. Lee, G. Hitt, E. Terada, J. H. Lee, Ethics of singing voice synthesis: Perceptions of users and developers, in: *Proc. International Society for Music Information Retrieval Conference*, 2022, pp. 733–740.
- [26] A. Swartz, Musicbrainz: A semantic web service, *IEEE Intelligent Systems* 17 (2002) 76–77.
- [27] A. Matamoros-Fernández, A. Rodríguez, P. Wikström, Humor that harms? examining racist audio-visual memetic media on tiktok during covid-19, *Media and Communication* 10 (2022) 180–191.
- [28] Reuters, Pop star kesha releases first single after label dispute, *Reuters* (2016). URL: <https://www.reuters.com/article/us-music-kesha-idUSKCN0XQ296>.
- [29] R. Brunner, Why is taylor swift re-rerecording her old albums?, 2021. URL: <https://time.com/5949979/why-taylor-swift-is-rerecording-old-albums/>.