# No Encore: Unlearning as Opt-Out in Music Generation

**Jinju Kim**[1,2*] **Taehan Kim**[2,3*] **Abdul Waheed**[2] **Jong Hwan Ko**[1] **Rita Singh**[2]
[1]Department of Electrical and Computer Engineering, Sungkyunkwan University
[2]Language Technologies Institute, Carnegie Mellon University [3]Sogang University
{jinjukim,abdulw,rsingh}@andrew.cmu.edu taehan@sogang.ac.kr jhko@skku.edu

## Abstract

AI music generation is rapidly emerging in the creative industries, enabling intuitive music generation from textual descriptions. However, these systems pose risks in exploitation of copyrighted creations, raising ethical and legal concerns. In this paper, we present preliminary results on the first application of machine unlearning techniques from an ongoing research to prevent inadvertent usage of creative content. Particularly, we explore existing methods in machine unlearning to a pre-trained Text-to-Music (TTM) baseline and analyze their efficacy in unlearning pre-trained datasets without harming model performance. Through our experiments, we provide insights into the challenges of applying unlearning in music generation, offering a foundational analysis for future works on the application of unlearning for music generative models.

## 1 Introduction

With remarkable breakthroughs in generative AI, high-quality content creation has become widely available in domains with only a precise design of text input prompt [1, 2, 3, 4, 5, 6]. While advances enable music creation [7], protection of creators' rights has become an urgent industry concern [8, 9]. Problems range from unauthorized use of copyrighted material, fairness in compensation, and infringement. With European Union's AI Act and General Data Protection Regulation (GDPR), these problems are further amplified [10]. Hence, major AI music generation systems and open data platforms are beginning to establish safeguards against unauthorized use of copyrighted material for AI training [11, 12, 13, 14]. Musicians have also voiced strong opposition to the use of copyrighted works for AI without permission [15, 16]. This reflects a growing trend: while the industry acknowledges AI's potential as a creative tool, it emphasizes that innovation must not come at the expense of artists' ownership and copyright protection.

As music generation service provider, it is crucial to ensure that artists who opt-out of are not exploited during pre-training and after model deployment. To address these challenges, machine unlearning (MU), a method designed to selectively remove the knowledge of certain datasets from a model's parameters can serve as a solution [17]. Instead of having to retrain a large model, MU ensures that information derived from specific datasets no longer affects the model's outputs with fine-tuning. MU has been explored in various domains of generative AI for its potential of removing unwanted dataset, concept, and output without harming generative performance [18, 19, 20, 21, 22, 23]. By integrating MU techniques into music AI, a solution for ethical and compliant music generation can be offered.

This study is the first to explore machine unlearning in Text-to-Music, establishing a foundation for future research on applying unlearning to musical content generation. Our experimental results reveal

*Work done while visiting Carnegie Mellon University.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Music.

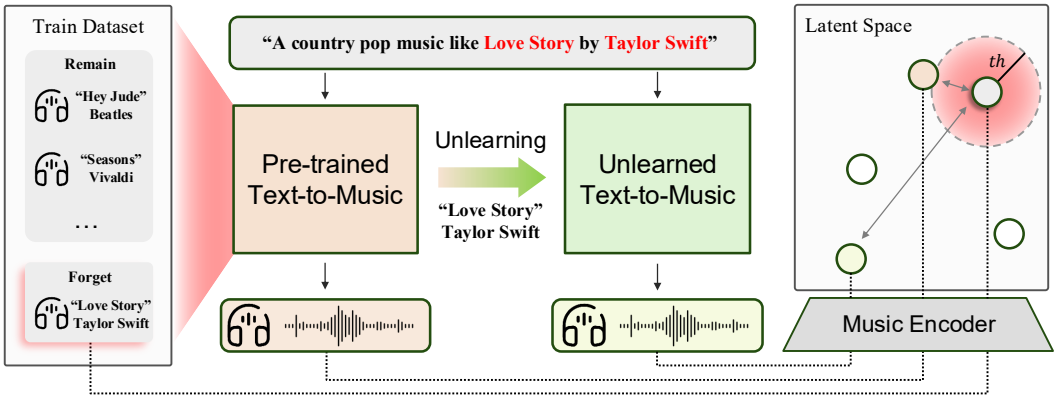


Figure 1: An overview of unlearning and its role in music generation: when a request is made to remove a specific piece of music from the pre-training dataset (the forget set), unlearning provides a potential solution. The goal is not only to erase the model's reliance on that data, but also to ensure that future generations diverge from the copyrighted material to ensure safe music generation.

the limitations of existing unlearning methods and analysis that point toward directions for safe music generation.

## 2 Related Works

**Text-to-Music Generation** Text-to-Music (TTM) models [24, 25, 26, 27, 28, 29] aim to generate music that aligns with provided textual descriptions, allowing users to specify musical characteristics such as genre, mood, instrumentation, and style. A common practice in developing these models involves training on large-scale corpora that often contain copyrighted material, frequently sourced from platforms like YouTube or other web repositories. This reliance on potentially sensitive and proprietary audio data raises ethical and legal questions.

**Generative Unlearning** Generative unlearning applies the principles of MU to generative AI to remove the influence of particular data or concepts from the generated distribution while preserving the model's ability to produce high-quality, diverse outputs [30]. Approaches often involve modifying or fine-tuning a pre-trained parameters to "forget" certain patterns without significantly degrading performance on the rest of the data distribution. These techniques have been explored in vision, language, and speech domains [31, 23]. Music, however, remains underexplored in machine unlearning. Our work represents a preliminary step toward extending generative unlearning frameworks to the complex and underexplored terrain of music generation. Please see Appendix B for extended related works.

## 3 Machine Unlearning for Music Generation

We investigate machine unlearning in the context of music generation with the goal of removing the influence of specific training subsets while preserving model performance. Our study is guided by the following question: *Can existing machine unlearning methods be directly applied to erase pre-trained musical knowledge in music generation models?*

### 3.1 Problem Formulation

In Figure 1, we consider a pre-trained TTM model parameterized by $\theta$, which maps a natural language prompt $x$ to a corresponding musical output $y$. Given a forget set $\mathcal{F} = (x^f, y^f)$ consisting of text-music pairs that should no longer influence the model's behavior, our goal is to modify the parameters of $\theta$ such that the model effectively "forgets" this data. Concretely, we require that:

- For a forget prompt $x^f$ that targets $y^f$ in the forget set, the updated model $\theta^-$ should no longer generate the corresponding music.

- For all other prompts $x \notin \mathcal{F}$, the updated model $\theta^-$ should preserve the original generation performance of $\theta$. We refer to these as remain set denoted by $x^r \in \mathcal{R}$.

### 3.2 Unlearning Methods

In this paper, we explore two representative machine unlearning approaches for TTM on MusicGen [24]. See Appendix A.1 for details on the model.

Following prior work [19], we implement **Gradient Ascent (GA)**, or Negative Gradient (NG) [19, 32, 20], which maximizes the training loss for $(x^f, y^f) \in \mathcal{F}$. By pushing $\theta$ in the direction opposite to standard training, GA attempts to revert learned influence from $\mathcal{F}$. We also implement **Random Labeling (RL)** which aims to break the learned associations of text prompts $x^f$ and corresponding target $y^f$. RL reassigns new random targets $\tilde{y}$ to $x^f$ and thus destroys the original correspondence of $(x^f, y^f)$. This forces the model to disregard the pre-trained alignment between text prompts and their associated musical outputs [30].

## 4 Experiments

### 4.1 Experimental Setups

**Backbone** Experiments are performed on MusicGen on the standard hyperparameter settings from the original paper [24]. MusicGen is pre-trained with 20k hours of licensed music comprised of internal dataset, ShutterStock[2], and Pond5[3] music data. Please refer to Appenndix A for details.

**Datasets** To remove pre-trained knowledge, we utilize approximately 500 samples from Pond5 dataset as the forget set. MusicCaps dataset [28] is employed as the remain set. Both datasets consist of extensive music-text pairs spanning a wide range of genres and styles. In our study, MusicCaps serves as $\mathcal{R}$ while Pond5 serves as $\mathcal{F}$. Specifically, we assess the preservation of performance on $\mathcal{R}$ to evaluate how unlearning with $\mathcal{F}$ affects the model's ability to generate text-coherent and diverse music. Please refer to Appendix A.2 for details.

**Evaluation Metrics** We assess how each unlearning methods affect the MusicGen in three key objectives, following benchmark of MusicCaps[28]; plausibility, alignment, and conceptual shifts. For plausibility, we adopt Fréchet Audio Distance (**FAD**) [33] using Tensorflow's VGGish model. For alignment, we evaluate using Kullback-Leibler Divergence (**KL**) with state-of-the-art audio classifier PaSST [34]. Lastly, conceptual shift is evaluated to assess how coherent the generated music remains with the text prompt using Contrastive Language-Audio Pretraining score (**CLAP**) [35].

### 4.2 Results

We present quantitative results to evaluate the impact of unlearning methods on retain and forget sets. Ideally, a well unlearned model should maintain its music generation performances on the remain set while selectively performing worse – failing to generate compliant music – when requested to generate audio with forget set.

Table 1: Quantitative results on Pond5 forget set. For the forget set, we expect degraded performance as denoted by arrows.

| Method | FAD (↑) | KL (↑) | CLAP (↓) |
|---|---|---|---|
| Original | 3.334 | **1.229** | 0.349 |
| Gradient Ascent (GA) | 3.866 | 1.158 | **0.351** |
| Random Labeling (RL) | **3.875** | 1.227 | 0.320 |

**Can existing machine unlearning methods erase pre-trained musical knowledge?** In Table 1, we evaluate models updated through unlearning on the Pond5 forget set. In terms of musical plausibility,

[2]https://www.shutterstock.com/

[3]www.pond5.com

FAD indicates that both GA and RL achieve partial forgetting: the music generated by unlearned models sounds less natural and exhibits lower audio quality. While unlearning is generally expected to reduce performance across all metrics, we observe the opposite trend for KL. This suggests that, although the unlearned models generate audio with degraded fidelity, their outputs remain similar to the forget set in terms of broad musical class (e.g., pop music). For CLAP scores, GA increases text–music alignment, whereas RL decreases it. GA appears unable to disentangle fidelity from semantic alignment. It primarily produces less natural music while still leaning on pre-trained knowledge. In contrast, RL maps prompts to essentially random outputs, teaching the model to ignore the text prompt. However, given the decrease in KL, we infer that RL only erases a limited set of anchor concepts while the model continues to retain some target knowledge.

Table 2: Quantitative results on MusicCaps remain set. For the remain set, an ideal unlearned model is desired to perform well. The scores are consistently lower on remain set because the model was never trained on this dataset, whose distribution differs from Pond5.

| Method | FAD (↓) | KL (↓) | CLAP (↑) |
|---|---|---|---|
| Original | **4.905** | **1.445** | **0.280** |
| Gradient Ascent (GA) | 9.458 | 1.777 | 0.226 |
| Random Labeling (RL) | 9.933 | 1.696 | 0.239 |

**Can existing machine unlearning methods preserve music generative performances?** In Table 2, we report results on the MusicCaps remain set. An ideal unlearned model should preserve its knowledge on the remain set, maintaining performance comparable to the original model. For unlearning methods, the FAD score increases substantially, indicating that GA and RL degrade the overall quality of generated music, particularly on distributions unseen during training. Furthermore, KL increases while CLAP decreases, suggesting that unlearning significantly diminishes both the fidelity and prompt alignment of the generated outputs.

## 5 Discussion

As we pioneer unexplored questions of unlearning in music, it is critical to reflect on the limitations of current approaches, especially in the context of musical copyrights and safe generative practices. Music is inherently complex, combining melody, harmony, rhythm, timbre, and performance, all of which interact in subtle and subjective ways. Determining whether a generated piece constitutes permissible stylistic inspiration or impermissible replication is far from straightforward.

For machine unlearning to be meaningfully applied in this domain, we must be able to distinguish which features of music should be retained (e.g., general compositional ability) and which should be removed (e.g., protected melodies, identifiable stylistic traits). We argue and seek progress in unlearning for music above collaboration with musical professionals and copyright experts to define the boundaries of creative influence and infringement. Any unlearning framework must navigate these domain-specific nuances to ensure that removing unwanted content does not diminish the model's creative potential, but instead supports a more responsible ecosystem for AI-assisted music creation.

## 6 Conclusion and Future Work

We present preliminary results of an ongoing study, the first systematic exploration of applying machine unlearning techniques to music generation. The motivation is centered on reflecting recent trends in AI copyright regulations and necessity of IP protection in music generation. We aimed to demonstrate the feasibility of removing knowledge of specific datasets from model parameters to prevent the generation of plagiarized music.

Our experiments indicate that existing machine unlearning methods are not effective when directly applied to Text-to-Music generation models. Unlearning results in model performance degradation for remain set in all evaluation measures. Additionally, our work highlights the complexity of problem formulation of unlearning in music generation. Clearer definitions of unlearning targets and precise measures to evaluate successful unlearning are required for future work. Overall, these findings

establish strong baselines and underscore key directions for continued research on applying machine unlearning for music.

## Acknowledgments and Disclosure of Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (RS-2024-00345732), and by the Institute of Information & Communications Technology Planning & Evaluation(IITP) under multiple grants funded by the Korea government(MSIT), including AI Star Fellowship Support Program(Sungkyunkwan University)(RS-2025-25442569), AI Excellence Global Innovative Leader Education Program (RS-2022-00143911), and AI Semiconductor Innovation Research Center (10692981). This work was also supported by the IITP-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT) (RS-2021-II212052).

## References

[1] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8821–8831. PMLR, 18–24 Jul 2021.

[2] OpenAI. Introducing chatgpt. OpenAI Product/Research Preview, nov 2022.

[3] Sanyuan Chen, Shujie Liu, Long Zhou, Yanqing Liu, Xu Tan, Jinyu Li, Sheng Zhao, Yao Qian, and Furu Wei. Vall-e 2: Neural codec language models are human parity zero-shot text to speech synthesizers. 2024.

[4] OpenAI. Sora: Creating video from text. OpenAI Product/Research Post, feb 2025. Accessed: 2025-08-30.

[5] Haven Kim, Zachary Novack, Weihan Xu, Julian McAuley, and Hao-Wen Dong. Video-guided text-to-music generation using public domain movie collections. *ISMIR 2025*, 2025.

[6] Keon Lee, Dong Won Kim, Jaehyeon Kim, Seungjun Chung, and Jaewoong Cho. DiTTo-TTS: Diffusion transformers for scalable text-to-speech without domain-specific factors. In *The Thirteenth International Conference on Learning Representations*, 2025.

[7] Chris Donahue, Shih-Lun Wu, Yewon Kim, Dave Carlton, Ryan Miyakawa, and John Thickstun. Hookpad Aria: A copilot for songwriters. In *Extended Abstracts for the Late-Breaking Demo Session of the 25th International Society for Music Information Retrieval Conference*, 2024.

[8] João Pedro Quintais. Generative ai, copyright and the ai act. *Computer Law & Security Review*, 56:106107, 2025.

[9] U.S. Copyright Office. Copyright and artificial intelligence, part 2: Copyrightability. Technical report, U.S. Copyright Office, jan 2025.

[10] European Parliament and Council of the European Union. Regulation (eu) 2016/679 (general data protection regulation). Official Journal of the European Union (L 119), pp. 1–88, April 27 2016. Accessed: 30 August 2025.

[11] Syed Irfan Ali Meerza, Lichao Sun, and Jian Liu. Harmonycloak: Making music unlearnable for generative ai. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 430–448, 2025.

[12] Sony Music Group. Declaration of ai training opt out. `https://www.sonymusic.com/sonymusic/declaration-of-ai-training-opt-out/`, May 2024.

[13] Warner Music Group. Wmg statement regarding ai technologies. PDF statement, July 2024. Accessed: 30 August 2025.

[14] The Wall Street Journal. Universal, warner and sony are negotiating ai licensing rights for music. Online article, June 2 2025. Accessed: 30 August 2025.

[15] Sam Tabahriti. Musicians release silent album to protest u.k.'s ai copyright changes. *Reuters*. Accessed: 30 August 2025.

[16] John Thornhill. Help is coming in the ai copyright wars. *Financial Times*. Accessed: 30 August 2025.

[17] Yinzhi Cao and Junfeng Yang. Towards making systems forget with machine unlearning. In *2015 IEEE Symposium on Security and Privacy*, pages 463–480, 2015.

[18] Rohit Gandikota, Joanna Materzynska, Jaden Fiotto-Kaufman, and David Bau. Erasing concepts from diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2426–2436, 2023.

[19] Anvith Thudi, Gabriel Deza, Varun Chandrasekaran, and Nicolas Papernot. Unrolling sgd: Understanding factors influencing machine unlearning. In *2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P)*, pages 303–319. IEEE, 2022.

[20] Kang Gu, Md Rafi Ur Rashid, Najrin Sultana, and Shagufta Mehnaz. Second-order information matters: Revisiting machine unlearning for large language models. *arXiv preprint arXiv:2403.10557*, 2024.

[21] Zhen Liu, Guangxu Dou, Zhenning Tan, Yanjun Tian, and Ming Jiang. Machine unlearning in generative ai: A survey. *arXiv preprint arXiv:2407.20516*, 2024.

[22] Chongyu Fan, Jiancheng Liu, Eric Wong Yihua Zhang, Dennis Wei, and Sijia Liu. Salun: Empowering machine unlearning via gradient-based weight saliency in both image classification and generation. *arXiv preprint arXiv:2310.12508*, 2024.

[23] Taesoo Kim, Jinju Kim, Dong Chan Kim, Jong Hwan Ko, and Gyeong-Moon Park. Do not mimic my voice : Speaker identity unlearning for zero-shot text-to-speech. In *Forty-second International Conference on Machine Learning*, 2025.

[24] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Defossez. Simple and controllable music generation. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 47704–47720. Curran Associates, Inc., 2023.

[25] Chang Li, Ruoyu Wang, Lijuan Liu, Jun Du, Yixuan Sun, Zilu Guo, Zhenrong Zhang, and Yuan Jiang. Qa-mdt: Quality-aware masked diffusion transformer for enhanced music generation, 2024.

[26] Zhengcong Fei, Mingyuan Fan, Changqian Yu, and Junshi Huang. Flux that plays music, 2024.

[27] Sanjoy Chowdhury, Sayan Nag, K J Joseph, Balaji Vasan Srinivasan, and Dinesh Manocha. Melfusion: Synthesizing music from image and language cues using diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26826–26835, June 2024.

[28] Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. Musiclm: Generating music from text, 2023.

[29] Zachary Novack, Ge Zhu, Jonah Casebeer, Julian McAuley, Taylor Berg-Kirkpatrick, and Nicholas J. Bryan. Presto! distilling steps and layers for accelerating music generation. In *International Conference on Learning Representations (ICLR)*, 2025.

[30] Aditya Golatkar, Alessandro Achille, and Stefano Soatto. Eternal sunshine of the spotless net: Selective forgetting in deep networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9304–9312, 2020.

[31] Juwon Seo, Sung-Hoon Lee, Tae-Young Lee, Seungjun Moon, and Gyeong-Moon Park. Generative unlearning for any identity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9151–9161, 2024.

[32] XiaoHua Feng, Chaochao Chen, Yuyuan Li, and Zibin Lin. Fine-grained pluggable gradient ascent for knowledge unlearning in language models. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 10141–10155, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[33] Kevin Kilgour, Mauricio Zuluaga, Dominik Roblek, and Matthew Sharifi. Fréchet audio distance: A metric for evaluating music enhancement algorithms, 2019.

[34] Khaled Koutini, Jan Schluter, Hamid Eghbal-Zadeh, and Gerhard Widmer. Efficient training of audio transformers with patchout. *Interspeech*, 2021.

[35] Benjamin Elizalde, Soham Deshmukh, Mahmoud Al Ismail, and Huaming Wang. Clap learning audio concepts from natural language supervision. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[36] Nicky Pochinkov and Nandi Schoots. Dissecting language models: Machine unlearning via selective pruning, 2024.

# A Experimental Details

## A.1 Model Configurations

**MusicGen** MusicGen is a single-stage autoregressive Transformer-based music generation model [24]. It operates directly over compressed discrete audio tokens—derived via an audio compression model—and is capable of producing high-quality music conditioned on text descriptions or melodic input. By using an efficient token interleaving strategy, MusicGen bypasses the need for multi-stage architectures or cascading models, enabling fast and controllable generation of mono or stereo audio in a single forward pass. We utilize the `MusicGen-small` variant under constrained computational resources.

Notably, MusicGen's training data primarily comprises audio purchased from royalty-free music platforms like Shutterstock[4] and Pond5[5]. This careful selection of legally licensed content distinguishes it from many TTM approaches that rely on web-crawled, potentially copyrighted music. As a result, MusicGen offers a suitable foundation for exploring unlearning strategies that aim to remove specific training influences while maintaining lawful and ethically sourced datasets.

**Hyperparameters** Our experiments are based on the standard hyperparameter settings from MusicGen, with specific adjustments to the learning rate and training duration to align with our unlearning objectives. We adopt a batch size of 6 without gradient accumulation, leveraging NVIDIA A100 (80GB) and RTX 4090 GPUs for training. To accommodate the smaller batch size and ensure stable optimization, we reduce the learning rate from $1 \times 10^{-3}$ to $1 \times 10^{-4}$.

**Unlearning Implementations** For the GA unlearning method, we conduct 10,000 training steps. For RL, we perform 1,000 training steps. We halt at these steps as further training drastically degrades model performance with explosive loss, referred to as catastrophical forgetting in the unlearning domain. This experimental setup enables us to effectively evaluate the impact of our unlearning strategies on both the target and unseen forget sets, ensuring that the model maintains its performance on the retained dataset despite the unlearning processes applied.

## A.2 Data Selection

**Forget Dataset** We utilize approximately 500 samples, representing 0.1% of Pond5's free trial dataset, for our unlearning experiments. The 500 samples were selected randomly using systematically designed crawling.

**Remain Dataset** For the remain set, we employ the MusicCaps dataset [28], which comprises extensive music-text pairs spanning a wide range of genres and styles. MusicCaps was utilized by MusicGen for evaluation purposes, providing a robust benchmark for assessing generative performance and text-audio alignment. In our study, MusicCaps serves as the remain set, allowing us to measure how well the model maintains its generative capabilities and alignment with textual prompts following the unlearning process. Specifically, we assess the preservation of performance on $\mathcal{R}$ to ensure that the unlearning of $\mathcal{F}$ does not adversely affect the model's ability to generate coherent and diverse music in response to various textual descriptions.

# B Related Works

## B.1 Generative Unlearning

Traditional unlearning methods strive for a theoretically grounded solution, attempting to remove the influence of $\mathcal{F}$ without extensive retraining for large models. A retrained model that requires a new training from scratch is referred to as Exactly Unlearned model. Many approaches seek closed-form parameter updates or leverage stored intermediate values from the original training process, aiming to restore parameters to a state equivalent to having never seen $\mathcal{F}$. While achieving true exactness is challenging, these methods represent a direction to approximate such parameter updates. For instance, methods such as Gradient Ascent attempt to reverse the impact of $\mathcal{F}$ by directly pushing the model's parameters away from the patterns learned from this data [19]. This involves maximizing the loss on $\mathcal{F}$, thereby reducing the model's reliance on it. Random Labeling (RL) aims to break the learned associations derived from $\mathcal{F}$ by randomly assigning new labels to those samples [30]. Techniques utilizing pruning remove or modify certain parameters or subnetworks within a model that are believed to encode the information related to $\mathcal{F}$ [36]. By systematically identifying and pruning weights associated with the forgotten data, the model's capacity to reproduce $\mathcal{F}$ content is reduced. Although pruning may affect overall capacity, when carefully applied, it can isolate and remove targeted knowledge while preserving essential functionality.

[4] https://www.shutterstock.com/

[5] https://www.pond5.com

In the vision domain, generative unlearning techniques have been explored to erase specific objects or styles from image synthesis models, ensuring that models no longer produce unwanted content while still generating plausible images [31, 22]. Similarly, in natural language processing, generative unlearning has been applied to large language models to remove sensitive textual patterns or biased ideas, allowing the models to remain fluent and coherent while mitigating undesirable language use [20]. While research remains relatively underexplored in the audio domain, a work to remove speaker identities from zero-shot text-to-music generation models have emerged [23].