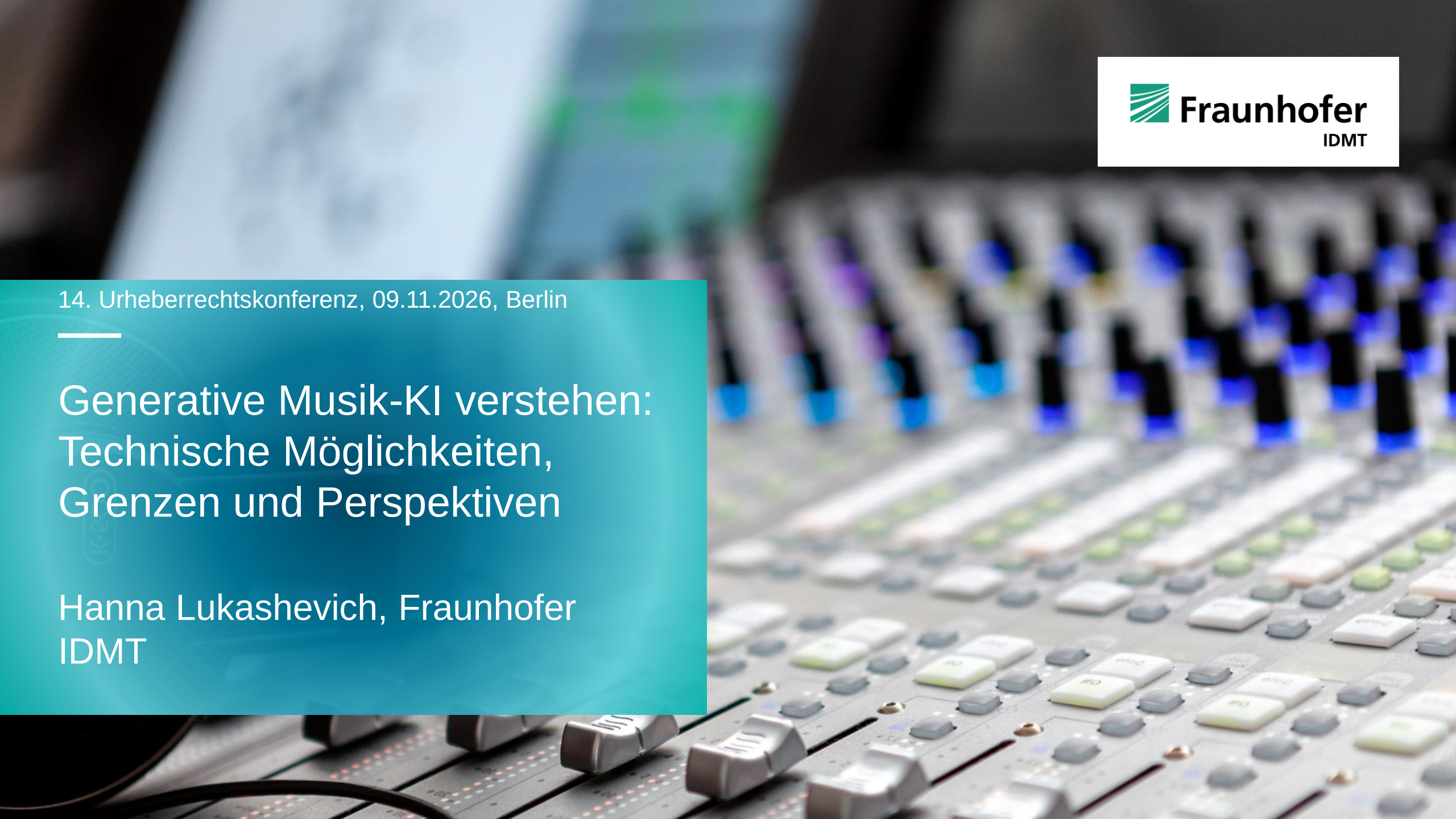


Fraunhofer IDMT

14. Urheberrechtskonferenz, 09.11.2026, Berlin

Generative Musik-KI verstehen: Technische Möglichkeiten, Grenzen und Perspektiven

Hanna Lukashevich, Fraunhofer IDMT

Fraunhofer-Gesellschaft
Auf einen Blick

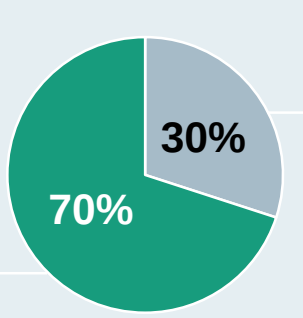
Anwendungsorientierte Forschung mit Fokus auf zukunftsrelevante Schlüsseltechnologien sowie auf die Verwertung der Ergebnisse in Wirtschaft und Industrie. Wegweiserin und Impulsgeberin für innovative Entwicklungen.

Rund 30 000
Mitarbeitende

74 Institute und Forschungseinrichtungen

3,6 Mrd. € Finanzvolumen
3,2 Mrd. € Vertragsforschung

Industrieaufträge und öffentlich finanzierte Forschungsprojekte



Grundfinanzierung durch Bund und Länder

 © Fraunhofer IDMT

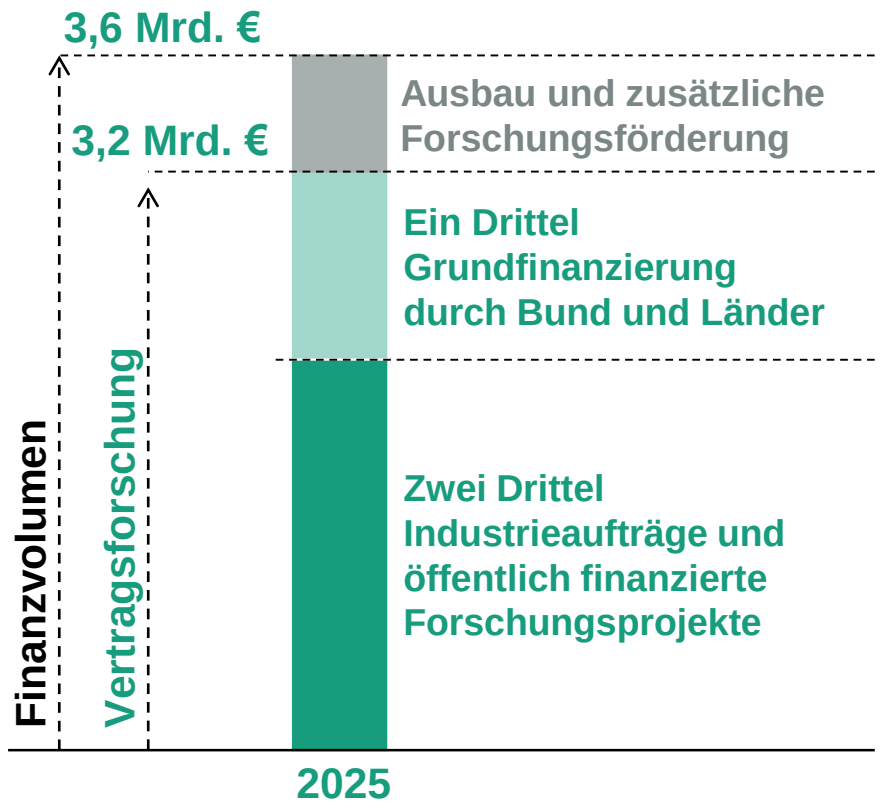
Fraunhofer IDMT

Fraunhofer-Gesellschaft

Auf einen Blick



● Hauptsitze der Fraunhofer-Institute und -Einrichtungen



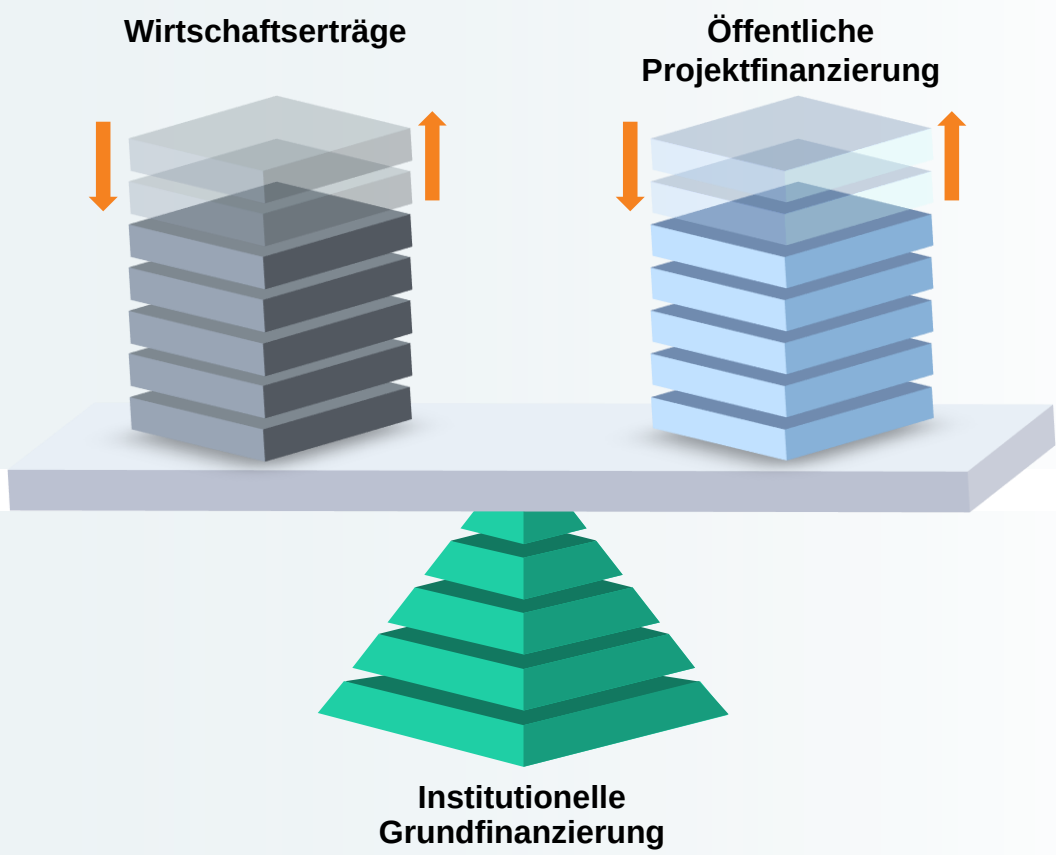
 © Fraunhofer IDMT **Public**

Fraunhofer IDMT

Fraunhofer-Gesellschaft

Im Fraunhofer-Modell werden mind. 2/3 der Erträge extern erwirtschaftet

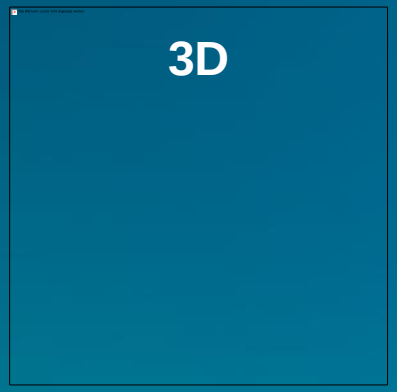
- Zusammen mindestens zwei Drittel der Erträge in der Vertragsforschung
- Ziel: **ausgeglichen**

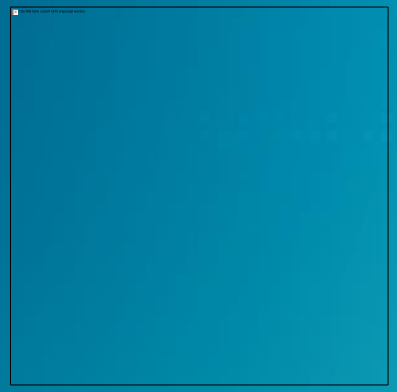


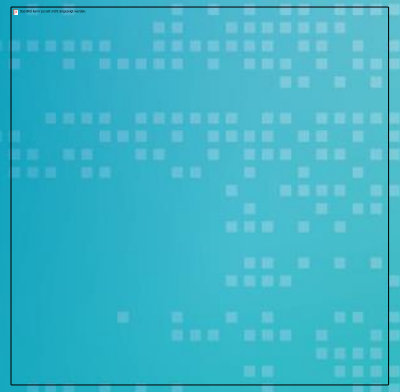
- Sicherung einer inflationsbereinigt stabilen, institutionellen Förderung
- Jährliche Steigerung von drei Prozent (Pakt für Forschung und Innovation)

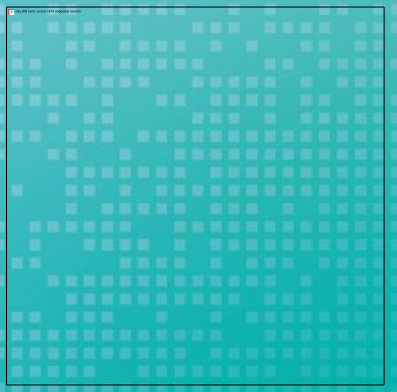
© Fraunhofer IDMT

Fraunhofer Institute for Digital Media Technology IDMT
„Where A.I. means audio intelligence.“









Smart Acoustics

Metadata & Music AI

Acoustic Monitoring

Media Security

© Fraunhofer IDMT

Über mich

Forschungsinteressen

- Musikanalyse
  - Tagging, Ähnlichkeitssuche, Transkription
- Audio-Matching
  - Audio-Fingerabdrücke, Inhaltsverfolgung
- **Generative KI für Musik**
  - **Kontrollierte Erzeugung, Anpassung**
  - **Detektion und Attribution**

@ Fraunhofer IDMT

- Seit 2006 am Institut mit Sitz in Ilmenau
- Head of Semantic Music Technologies seit 2014
- Head of Semantic Media Technologies seit 2026
  - Interdisziplinäres Team, auch Bild-/Videoanalyse
- Strategische Entwicklung
- Akquisition und Leitung von F&E-Projekten



 © Fraunhofer IDMT Public

Fraunhofer IDMT

Musikquiz: Mensch oder Maschine?





#1		Suno, Pop
#2		Suno, Rock
#3		Suno, Country

© Fraunhofer IDMT

Grenzen des Publikums

Deezer: **täglich** werden ca. **90.000 vollständig KI-generierte Titel** hochgeladen
> 50 % aller täglich angelieferten Songs

https://newsroom-deezer.com/2026/10/ai-music-detection-time-best-inventions-2026/
https://newsroom-deezer.com/2025/11/deezer-ipsos-survey-ai-music/

97 % konnten in einem **Blindtest** mit zwei KI-Songs und einem echten Song **keinen Unterschied** zwischen vollständig KI-generierter Musik und von Menschen gemachter Musik **feststellen**

52 % empfanden es als **unangenehm**, dass sie den Unterschied **nicht feststellen konnten**


DEEZER NEWSROOM

Deezer/Ipsos survey: 97% of people can't tell the difference between fully AI-generated and human made music – clear desire for transparency and fairness for artists

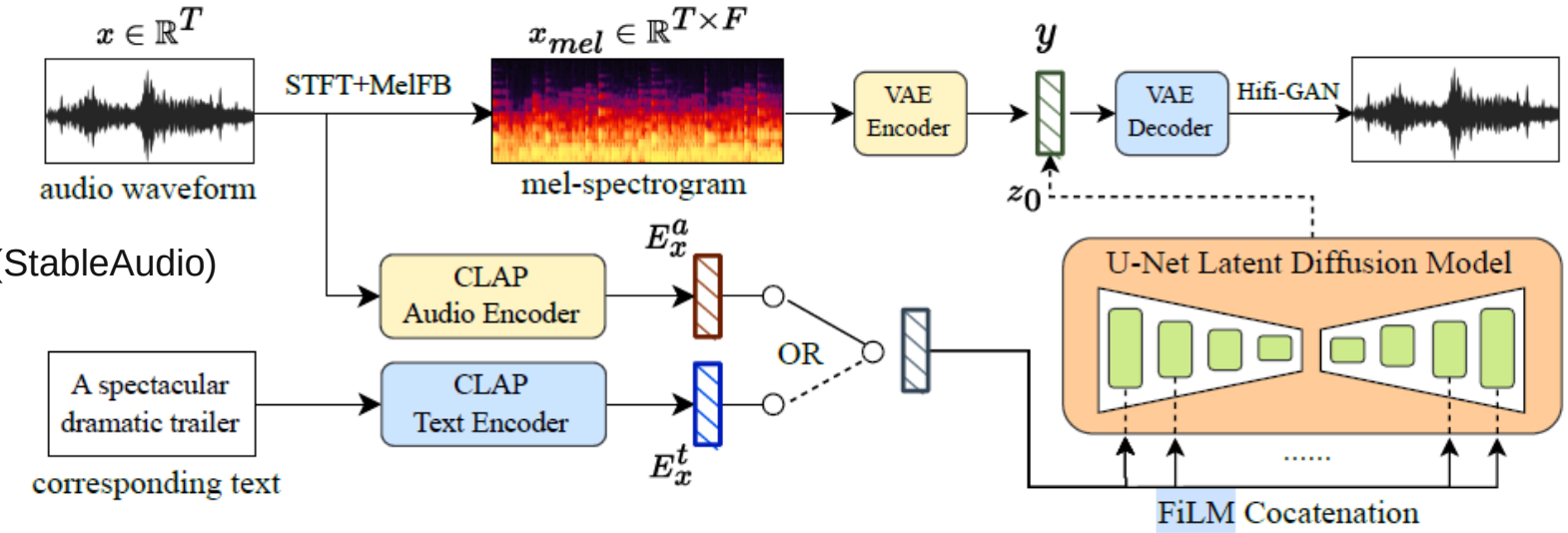
Deezer - Nov 12, 2025 • 9 min read

© Fraunhofer IDMT

Text-to-Audio-Modelle

- Mehrere Architektur-Tendenzen, z. B.
  - Autoregressive LM-basierte Modelle (MusicGen)
  - Nichtautoregressive, diffusionsbasierte Modelle (StableAudio)

MusicLDM

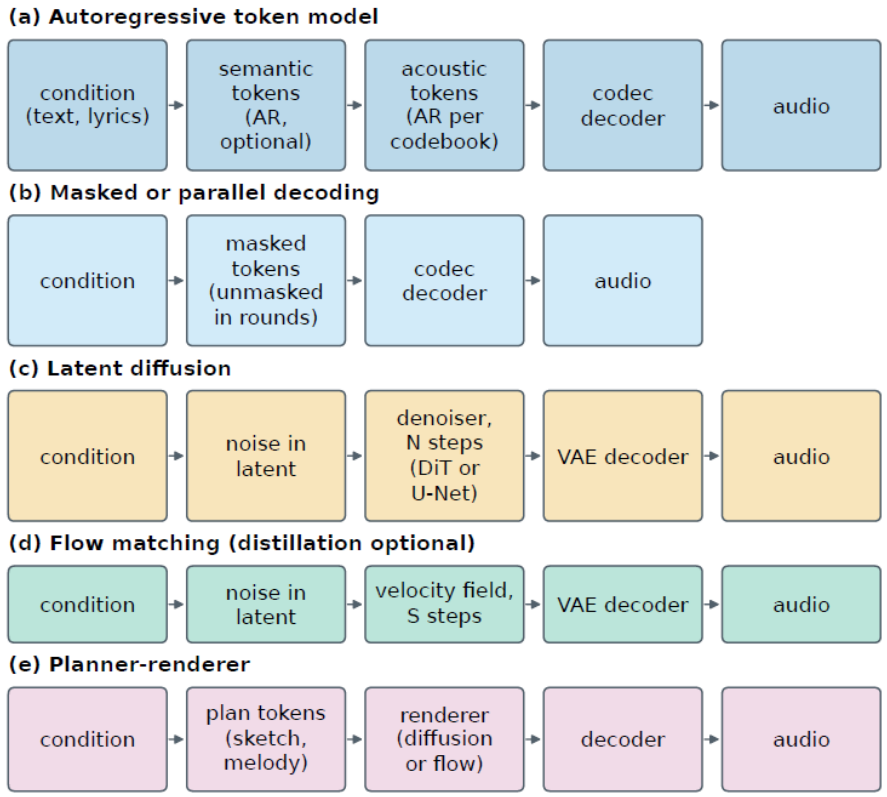


- Kombination aus 6 neuronalen Netzen
- CLAP (Contrastive Language-Audio Pretraining)
  - Foundation Model, Übereinstimmung von Text und Musik
  - Trainingsablauf: zuerst Self-supervised, danach Supervised
- VAE (Variational Autoencoder, Encoder-Decoder-Architektur)
  - Generiert eine interne (latente) Darstellung eines Spektrogramms und danach ein neues Spektrogramm
  - Trainingsablauf: Self-supervised
- U-Net LDF (Latent Diffusion Model)
  - Generiert einen Vektor zur Steuerung des VAE
- HiFi-GAN (Generative Adversarial Network)
  - Generiert die Audio-Wellenform

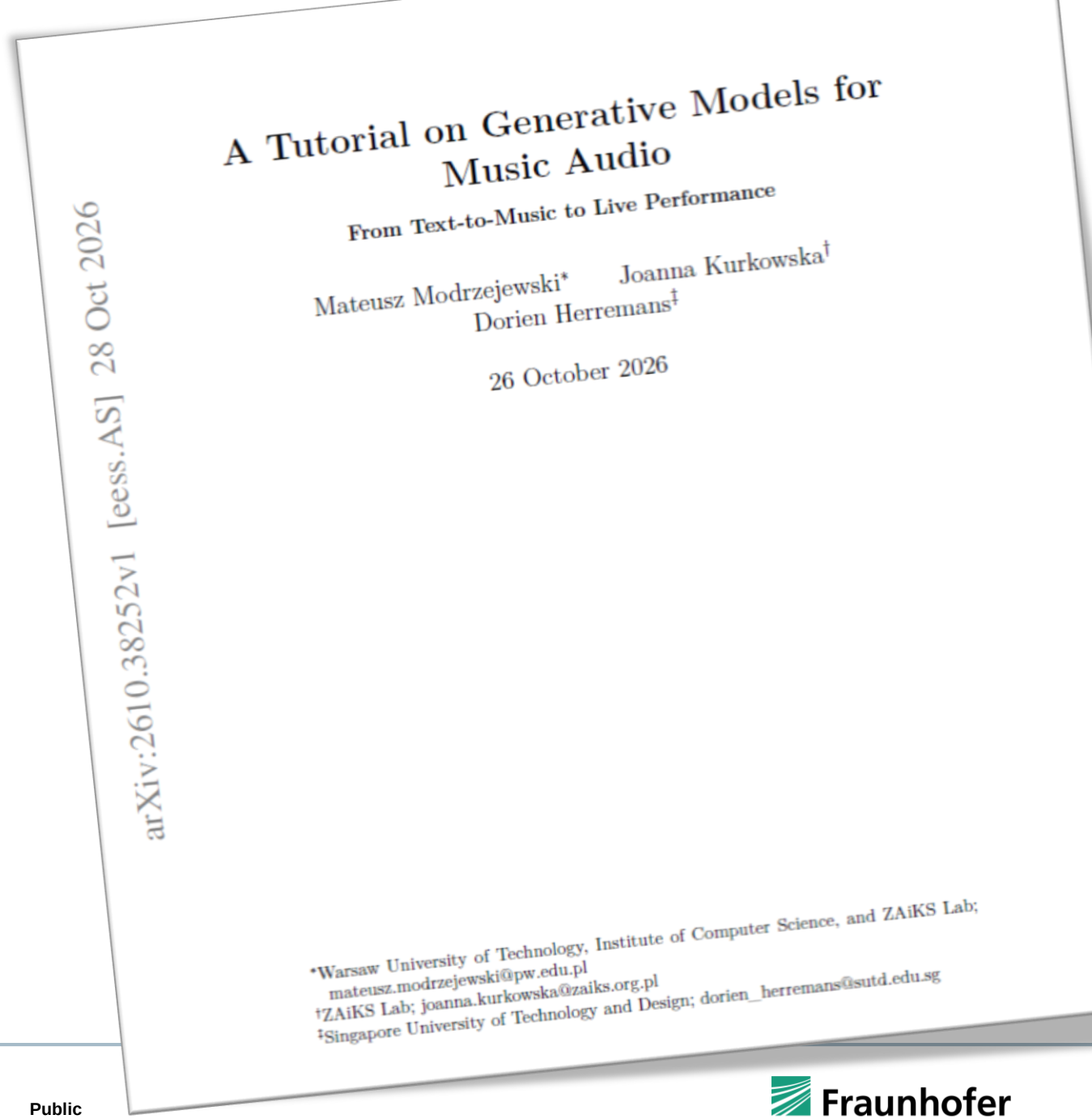
Source: Chen et al. (2024). MusicLDM: Enhancing novelty in text-to-music generation using beat-synchronous mixup strategies. In Proceedings of ICASSP, https://arxiv.org/pdf/2308.01546

© Fraunhofer IDMT Public

Tutorial on Generative Models

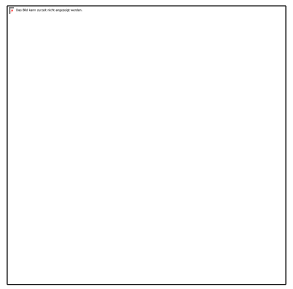


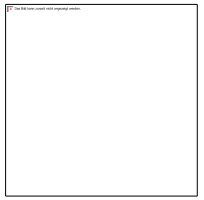
Modrzejewski, M., Kurkowska, J., & Herremans, D. (2026).
A Tutorial on Generative Models for Music Audio. arXiv preprint arXiv:2610.38252

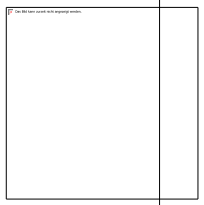


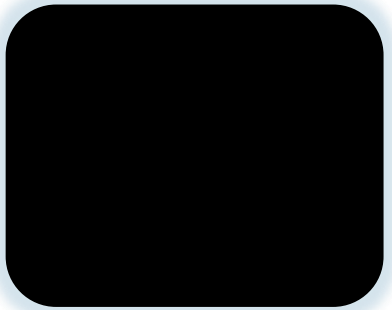
© Fraunhofer IDMT

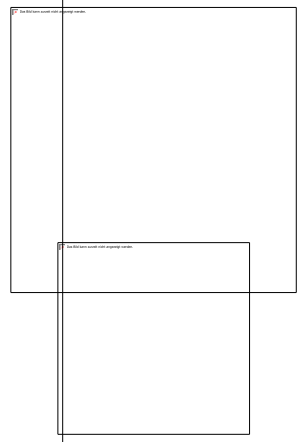
Grenzen des Trainingsmaterials











1 Million Songs

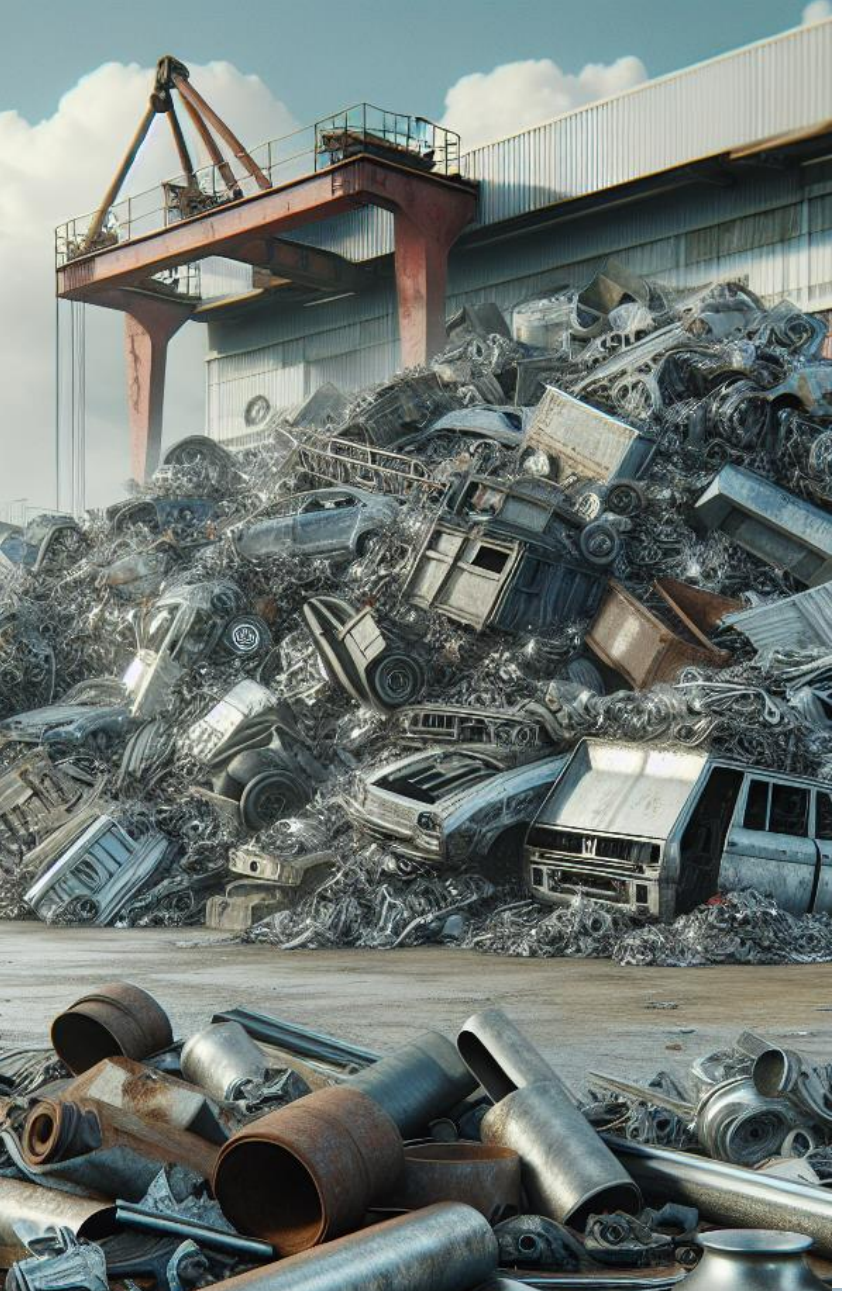
in (z. B. 20–100) Teile zerlegt *

zufällig in Gruppen kombiniert

Millionenfach durch verschiedene Schichten in mehreren neuronalen Netzen iteriert

um eine unbegrenzte Anzahl neuer Musikstücke auszugeben

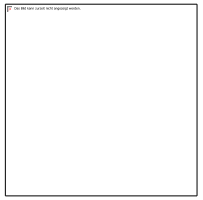
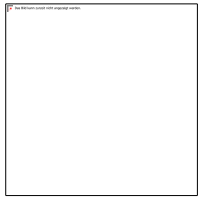
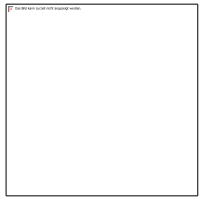
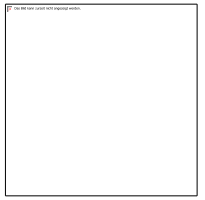
** ca. 43.000.000.000





Bilder: DALL·E 3

Grenzen des Trainingsmaterials: Vom Wertstoffhof zu Messer, Gabel, Löffel…

Altmetall	Revenue	Sortieren	Schmelzen	Business	Neue Produkte
					 …
Songs	**Revenue**	**KI-Training**	**Musikmaschine**	**Business**	**Neue Songs**

Virtuelle Produkte = x-mal Erlöse

© Fraunhofer IDMT

Grenzen der Zuordnung des KI-generierten Outputs



Nur Versprechen
ohne Nachweis, bisher …



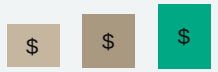
Nur Theorie: Wenn das **Trainingsmaterial** bekannt ist **UND** das **GenAI-Modell** einschließlich aller Gewichte usw. **bekannt** ist, besteht die **Möglichkeit**, eine **Verbindung** zwischen der generierten **KI-Ausgabe** und der Wirkung bestimmter Teile des **Trainingsmaterials** herzustellen ... allerdings hängt dies noch von vielen (unbekannten) Anforderungen ab. *

** Park et al. (2023). Trak: Attributing model behavior at scale. arXiv preprint arXiv:2303.14186. https://arxiv.org/abs/2303.14186*

© Fraunhofer IDMT

Wie kann man KI-Einnahmen an die Künstler verteilen, deren Musik für das KI-Training genutzt wird?

KI-Dienste erzielen Einnahmen



Labels und Verlage sammeln die Einnahmen

(Beispiel)

8 Milliarden Dollar

KI-Musikmarkt, 2030

Studie von GEMA und SACEM

Goldmedia für GEMA / SACEM (2024)

Labels und Verlage müssen die KI-Einnahmen zuweisen. Aber wer bekommt wie viel?

Künstler, deren Musik für das KI-Training verwendet wird



 © Fraunhofer IDMT

Creative Weight Attribution

Nachvollziehbare Verteilung von KI-Musikerlösen

- Faire Vergütung von Musik, die als KI-Trainingsmaterial dient
- Musikalische Merkmale analysieren
  - Bewertung nach Komplexität, Einzigartigkeit & Neuartigkeit
  - Musikologische Merkmale zu Melodie, Harmonie, Rhythmus
  - Kombination mit Output Similarity & Prompt Reference
- Gewichtung gemeinsam festlegen
  - Die Beteiligten vereinbaren, wie stark die verschiedenen musikalischen Einflüsse bei der Zuordnung der Erlöse zählen
  - Master- und Publishing-Rechte gezielt vergütbar
- Praxiserprobung
  - Ein erfolgreicher, bezahlter Pilot mit der niederländischen Verwertungsgesellschaft Buma/Stemra
  - Weitere Pilotprojekte folgen



DOI: 10.71568/daga2026.240

Attribution for Musical Generative AI: A New Perspective

Andrew McLeod[1], Peer-Uli May[2], Thomas Rekittke[2], Hanna Lukashevich[1]
[1] Fraunhofer Institute for Digital Media Technology IDMT, 98693 Ilmenau, Germany,
Emails: andrew.mcleod@idmt.fraunhofer.de, hanna.lukashevich@idmt.fraunhofer.de
[2] Allrights AIxchange UG (haftungsbeschränkt), 10997 Berlin, Germany,
Emails: pu@allrights-aixchange.com, tr@allrights-aixchange.com

Introduction

As musical generative AI systems have become more prevalent in recent years, questions have arisen about data provenance. Given the recent EU AI Act [2], revenue attribution methods—which distribute license revenues back to the original rights holders of the training data—have become an important part of the ecosystem. Given the projected music generative AI revenues—nearly $8 billion in 2030, according to a projection by GEMA and SACEM, two major European music rights organizations [1]—such methods will soon become an essential component of the musical ecosystem.

Generative AI companies seem to be happy to simply license training data in bulk from labels and collective management organisations (CMOs), leaving them to perform the attribution themselves. However, exactly how to fairly perform such attribution is unclear. In this paper, we analyze this problem of Gen AI Attribution, presenting in particular five prerequisites that a solution must fulfill for it to be practically feasible and accepted by both the music industry and generative AI companies. We further describe existing approaches to the problem and their fulfillment of the prerequisites.

We end by presenting Create Weight Attribution, our proposed solution, derived from Music Information Retrieval (MIR) principles, which is fair, transparent, and fulfills all of the described goals and requirements.

Solution Prerequisites

We divide the solution prerequisites into two categories. First, goals: These are features of a solution that we believe will lead to a healthier and more thriving music industry community.

(2) Emphasis on creativity. This goal is more aspirational in terms of the music industry as a whole. Musical development thrives on creativity and novelty, and a solution that simply emphasizes popularity could lead to a stagnation of new musical ideas. Instead, a solution that explicitly emphasizes creativity and novelty encourages the diversity of music, leading to a more interesting musical landscape, and ensuring that musicians who try something new are also rewarded.

Requirements

(3) No model or training access. Some existing approaches (especially academic approaches) require knowledge of—or even control over—the model or the model training process. For generative AI companies, however, such access and control is untenable. Their model design and training process remain valuable trade secrets, highly optimized for output quality and low cost. As such, any proposed solution must not require access to or control over either the model or the training process.

(4) Model-agnostic. Related to point (3), any solution must be agnostic to the specific type of model used. Although the specifics are unknown, generative AI companies almost certainly use a variety of model architectures and combinations thereof. Therefore, any solution designed to work with only a single type of architecture (e.g., feed-forward networks or Generative Adversarial Networks (GANs)) is unsuitable—even if no direct access or control is required.

(5) Separate master and publishing rights. This final requirement is less technical than the first two, relating instead to the music rights landscape. A song's owner

McLeod, A., May, P.-U., Rekittke, T., & Lukashevich, H. (2026). Attribution for Musical Generative AI: A New Perspective. Jahrestagung für Akustik 2026.

aixchange.

Fraunhofer IDMT

Möglichkeiten

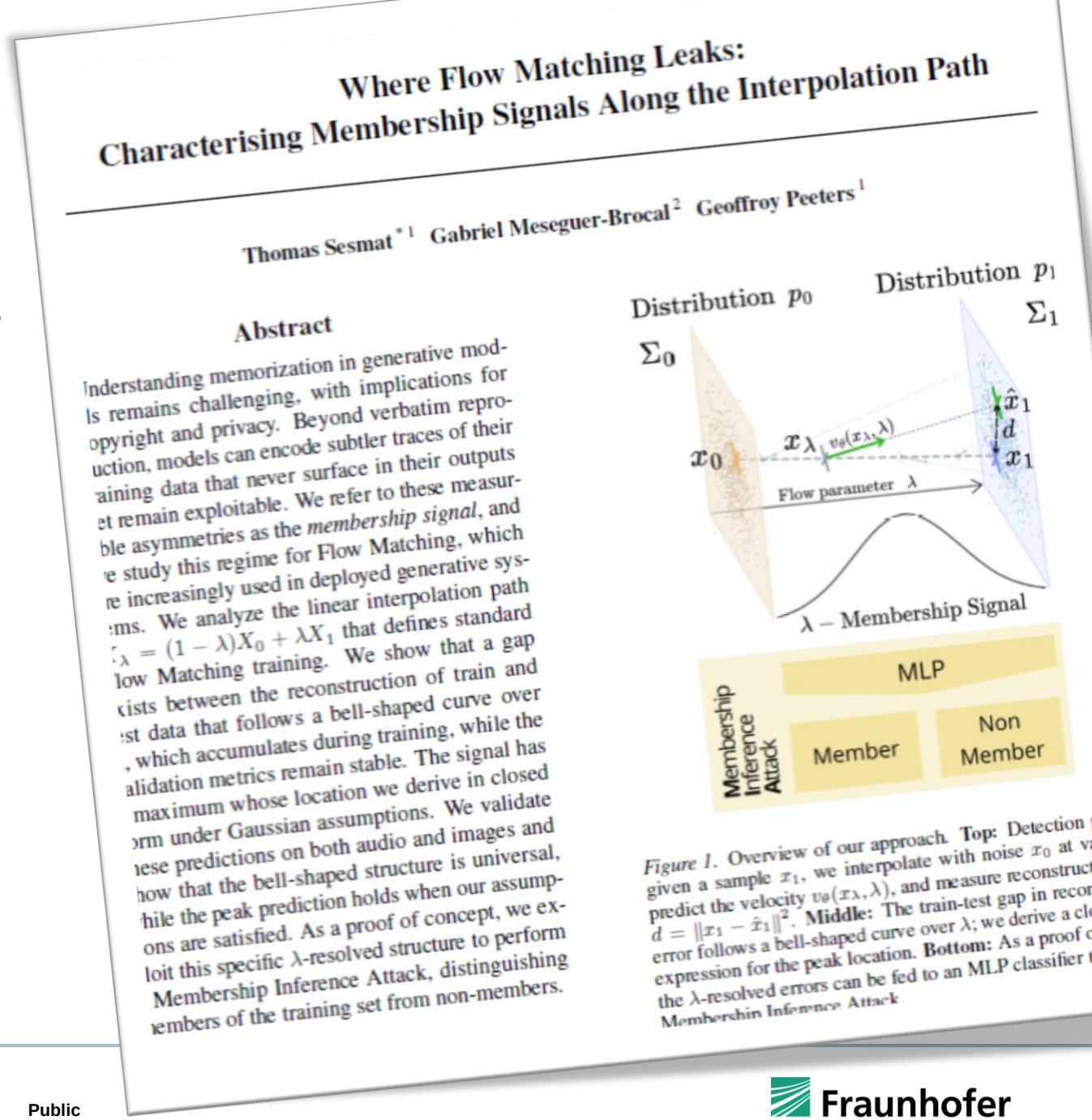
PoC: Trainingsdatenzugehörigkeit erkennen

- Ist ein bekannter Musiktitel wahrscheinlich Trainingsbestandteil?
  - Die Reproduzierbarkeit des Titels wird nicht untersucht
- Nutzt unterschiedliche Rekonstruktionsfehler von Trainings- und Nichttrainingsbeispielen bei Flow-Matching-Architekturen
- Bei 5 % False Positives werden je nach Datensatz 19,0 % bis 56,7 % der Trainingsbeispiele erkannt

Voraussetzungen: praktische Ableitung

- Interner Zugriff auf das Flow-Modell, eine Audio-API reicht nicht
- Nur für Flow-Matching-Architekturen
- Erfordert eine identische Vorverarbeitung
  - einschließlich des Audioencoders und des Latentraums
- Ein statistischer Hinweis, kein forensischer Nachweis
- Negative Ergebnisse schließen eine Trainingsnutzung nicht aus

Sesmat, T., Meseguer-Brocal, G., & Peeters, G. (2026). Where flow matching leaks: Characterising the membership signals along the interpolation path. In Proceedings of ICML.


Where Flow Matching Leaks:
Characterising Membership Signals Along the Interpolation Path

Thomas Sesmat [*1] Gabriel Meseguer-Brocal [2] Geoffroy Peeters [1]

Abstract

Distribution p_0
Distribution p_1
Σ_0
Σ_1
x_0
x_λ
$v_\theta(x_\lambda, \lambda)$
$\hat{x}_1$
d
x_1
Flow parameter λ
λ – Membership Signal
MLP
Membership Inference Attack
Member
Non Member

Figure 1. Overview of our approach. **Top:** Detection
given a sample x_1, we interpolate with noise x_0 at v
predict the velocity $v_\theta(x_\lambda, \lambda)$, and measure reconstruct
$d = \|x_1 - \hat{x}_1\|^2$. **Middle:** The train-test gap in recon
error follows a bell-shaped curve over λ; we derive a cl
expression for the peak location. **Bottom:** As a proof o
the λ-resolved errors can be fed to an MLP classifier
Membership Inference Attack

© Fraunhofer IDMT

Möglichkeiten

KI-generierte Inhalte erkennen

- Deconvolution-/Upsampling-Schichten der Generatoren hinterlassen charakteristische Peaks im Frequenzspektrum
- Muster ist vor allem von der Modellarchitektur abhängig
  - nicht von Trainingsdaten oder Gewichten
  - Technischer „Fingerprint“ der Generatorarchitektur
- Interpretierbare Alternative zu Black-Box-Detektoren
- 99 % Erkennung in mehreren getesteten Szenarien

Einschränkungen

- Keine universelle Signatur für alle GenAI-Architekturen
- Schwache Übertragbarkeit bei Architekturwechseln
- Anfällig für Resampling, Pitch-Shifting und andere Audiomanipulationen
- Ein Katz-und-Maus-Spiel, ähnlich wie bei Anti-Spam-Filtern

D. Afchar, G. Meseguer-Brocal, K. Akesbi, and R. Hennequin (2025), A Fourier Explanation of AI-music Artifacts, in Proceedings of the ISMIR 2025 Conference.



A FOURIER EXPLANATION OF AI-MUSIC ARTIFACTS

Darius Afchar Gabriel Meseguer-Brocal Kamil Akesbi Romain Her
Deezer Research, Paris, France
research@deezer.com

ABSTRACT

The rapid rise of generative AI has transformed music creation, with millions of users engaging in AI-generated music. Despite its popularity, concerns regarding copyright infringement, job displacement, and ethical implications have led to growing scrutiny and legal challenges. In parallel, AI-detection services have emerged, yet these systems remain largely opaque and privately controlled, mirroring the very issues they aim to address. This paper explores the fundamental properties of synthetic content and how it can be detected. Specifically, we analyze deconvolution modules commonly used in generative models and mathematically prove that their outputs exhibit systematic frequency artifacts – manifesting as small yet distinctive spectral peaks. This phenomenon, related to the well-known checkerboard artifact, is shown to be inherent to a chosen model architecture rather than a consequence of training data or model weights. We validate our theoretical findings through extensive experiments on open-source models, as well as commercial AI-music generators such as Suno and Udio. We use these insights to propose a simple and interpretable detection criterion for AI-generated music. Despite its simplicity, our method achieves detection accuracy on par with deep learning-based approaches, surpassing 99% accuracy on several scenarios

is estimated that 24% of musicians' revenues a
the next few years [13].

Most of these models are black box and ow
vate companies. There is a lack of regulation
velopment and use. In parallel to this rise of
detection services have started to emerge [14].
less, these services are not faring better than Ge
they are also mostly black box and owned by
ties. This begs the question: *What is specific ab*
Why do AI-detectors work? What cues may the

This paper may be seen as an interpretability
take a step back from the arms race between
AI-detectors and try to ask what is specific abou
content and how we might detect it. We analyz
lution modules that are common in generative
mathematically prove that their output will pr
frequency artifacts – *i. e.*, peaks (see Figure 3)
is related to the well-known *checkerboard*
Unlike previous empirical observations, we p
planation for its origin using Fourier's fran
tion 3). Interestingly, we prove these artifac
on the training data nor learned weights of ge
els but on their chosen architecture. In our ex
confirm the presence of this effect on multip
models and that their structure aligns with

Fraunhofer IDMT

Möglichkeiten

PoC: Hybride Inhalte erkennen

- „Wie viel KI steckt im Track?“ statt nur KI / Nicht-KI
- Schätzung eines kontinuierlichen KI-Energieanteils (0–100 %)
- Analyse von spektralen Artefakten neuronaler Audio-Decoder
- Training mit hybriden Mischungen aus menschlichen und KI-Stems
- Im Experiment: Ø 7,6 Prozentpunkte Fehler

Einschränkungen

- Bisher nur EnCodec-Rekonstruktionen
  - keine echten generierten Stems
- Erkennung stark instrumentabhängig
- Übertragbarkeit auf Suno, Udio etc. offen
- Einfluss von EQ, Kompression und Postprocessing noch nicht untersucht
- Proof of Concept, noch kein produktionsreifes Verfahren

de la Cruz, F. G., López-Ayala, D., Zinemanas, P., Molina, E., & Rocamora, M. (2026). How much AI is in this track? Quantifying the proportion of AI-generated stems in hybrid music mixtures. In Proceedings of the ISMIR 2026 Conference.



HOW MUCH AI IS IN THIS TRACK? QUANTIFYING THE PROPORTI
OF AI-GENERATED STEMS IN HYBRID MUSIC MIXTURES

Fernando Garcia de la Cruz[1] David López-Ayala[1] Pablo Zinemanas[2]
Emilio Molina[2] Martín Rocamora[1]
[1] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[2] BMAT Licensing S.L., Barcelona, Spain
fergarciadlc@gmail.com

arXiv:2510.17827v1 [eess.AS] 19 Oct 2025

ABSTRACT

AI-generated music is increasingly used at the stem level, with producers integrating synthetic drums, basslines, or vocals alongside human-performed instruments. However, current AI music detection systems are binary, treating tracks as either fully AI or fully human. In this paper, we reformulate AI music detection as a regression problem on a continuous AI energy ratio, $\alpha \in [0, 1]$. We propose a methodology that leverages a multi-track music dataset to assemble mixtures of human-performed and AI-reconstructed stems (obtained using a neural audio codec) with known proportions of each content type. Using this approach, we first show that a CNN-based model trained on fully AI-generated or human-performed tracks, which achieves $> 99\%$ accuracy as a binary detector, when faced with mixed content, yields an output that rises with the AI stems' energy contribution, acting as a noisy and miscalibrated estimator. Our analysis of the influence of different stems shows that detection sensitivity depends on the instrument and reflects its frequency content: drums and guitar carry strong codec-artifact signatures, while vocals and bass are less detectable. Based on these insights, we train a similar CNN-based model for regression of α, achieving MAE $= 0.076$ and $R^2 = 0.85$ on held-out mixtures from the same pipeline. These results suggest that the regression formulation is an initial promising step towards AI-music detection in realistic music production workflows for artists and highlights the need for automated det tools that are also aligned with disclosure obligation as those in Article 50 of the EU AI Act [3].

Current AI music detection systems achieve perfect accuracy (>99%) when classifying complete as fully AI-generated or human-produced [4, 5]. detectors exploit architectural artifacts: the charact spectral peaks introduced by transposed convolution in neural audio codec decoders [6]. Yet, it is uncle this binary paradigm applies when AI-generated ste combined with human performances on a single trac

In this work, we reformulate AI music detectio regression problem on a continuous AI energy rati $[0, 1]$, representing the proportion of a mix's total tic energy contributed by AI-generated stems. St this reformulation requires controlled mixtures of performed and AI-generated stems with known α v

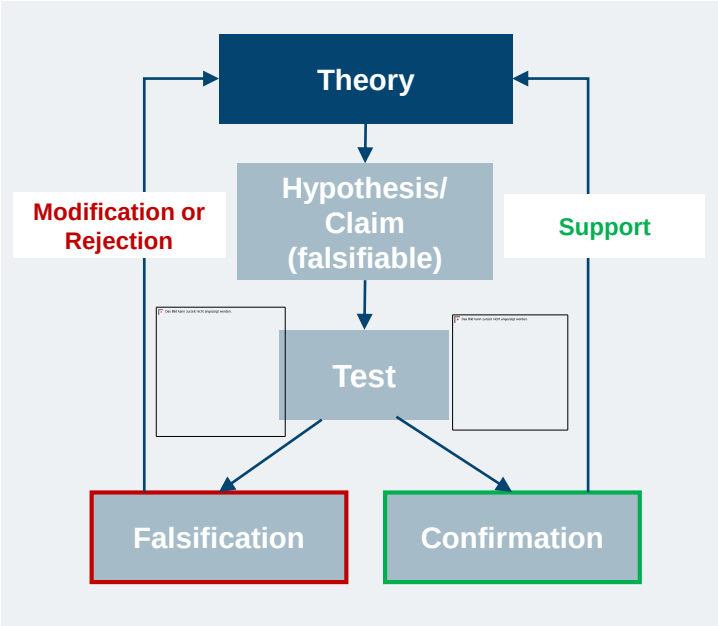
To enable this, we propose a methodology for co ing hybrid mixtures from multi-track music datase on three components: (i) encode–decode recons of human-performed stems through a neural audi which imprints the same architectural artifacts th detectors learn to detect while preserving the ori sical content; (ii) combinatorial stem replaceme for a track with n stems produces all 2^n real/AI tions, ensuring dense sampling of α; and (iii) ev the level of the AI energy ratio α, which correct

Fraunhofer IDMT

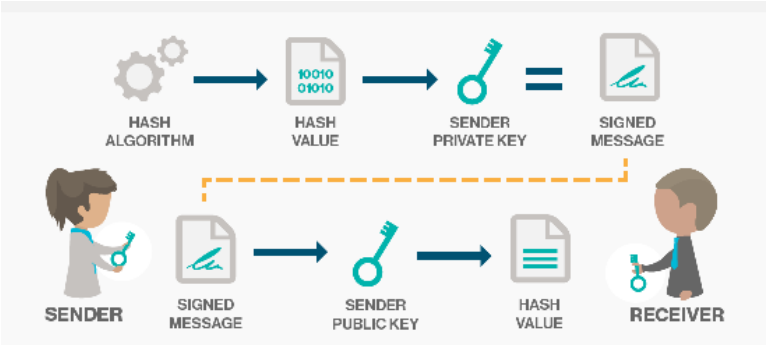
Von Aussagen zu Evidenz

Zwei komplementäre Strategien: Content Verification und Content Authentication

Verifikation: Prüfung von Aussagen über Inhalte
(via Signalanalyse, KI, statistische Analyse)



Authentifizierung: Herkunft und Prozess absichern
(digitale Signaturen, Hashing, Wasserzeichen)



Quelle: https://security.stackexchange.com/questions/87325/if-the-public-key-cant-be-used-for-decrypting-something-encrypted-by-the-priva



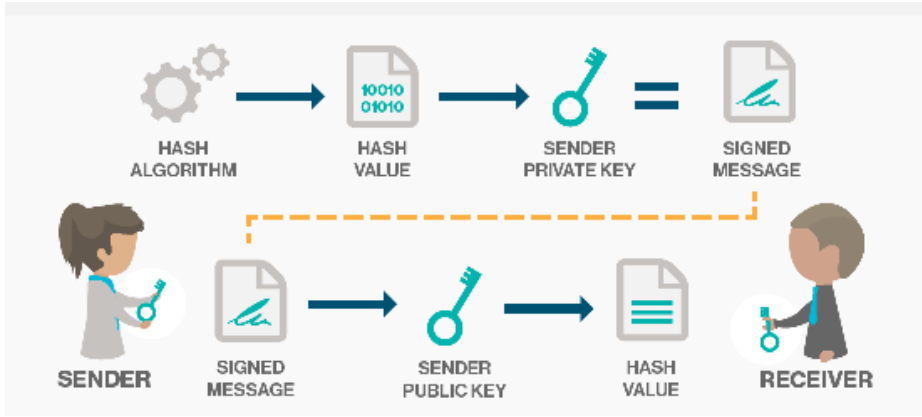
© Fraunhofer IDMT

Von Aussagen zu Evidenz

Authentifizierung: Belegbare Provenienz

- Grundprinzip: **Herkunft und Entstehungsprozess kryptografisch dokumentieren** und an den Inhalt binden
- Macht Aussagen zu **Herkunft, Integrität und Prozessschritten** belegbar
- Technische Prüfung: **Manifest lesbar? Signatur gültig? Vertrauenskette gültig? Content-Bindung intakt?**
- Einsetzbar in **kontrollierten Workflows**
- Kann **Transparenz- und Kennzeichnungspflichten unterstützen**, z. B. Art. 50 KI-VO
- **Ergänzt Content Verification, ersetzt sie nicht**





Quelle: https://security.stackexchange.com/questions/87325/if-the-public-key-cant-be-used-for-decrypting-something-encrypted-by-the-priva



© Fraunhofer IDMT

Zusammenfassung

Technische Möglichkeiten, Grenzen und Perspektiven

- **Generative Musik-KI eröffnet neue Produktionsmöglichkeiten.** Modelle erzeugen Audiosignale aus gelernten Mustern.
- **Detektion bleibt begrenzt.** KI-Musik, hybride Anteile und Trainingsnutzung erfordern unterschiedliche Prüfverfahren. Ein universeller Detektor fehlt.
- **Attribution braucht Bewertungsregeln.** Technische Ähnlichkeit allein bestimmt weder Urheberschaft noch eine faire Erlösverteilung.
- **Belastbare Entscheidungen brauchen mehrere Grundlagen.** Signalanalyse, dokumentierte Herkunft und gemeinsame Evaluation ergänzen sich. Unsicherheiten bleiben sichtbar.



Fraunhofer IDMT

Ausblick: Transparenz und Verantwortung im Musikökosystem

Offene Fragen für Forschung, Praxis und faire Vergütungsmodelle

Offene Gestaltungsfragen

Anreize zur Offenlegung

Wer legt KI-Anteile freiwillig offen, wenn dadurch die Vergütung sinken könnte?

Menschliche Beiträge in hybrider Musik

Wie lässt sich kreative Leistung jenseits des technisch geschätzten KI-Audioanteils bewerten?

Zuständigkeit im Ökosystem

Wer prüft Angaben und trägt Verantwortung: Kreative, Distributoren, Plattformen oder Verwertungsgesellschaften?

Kontrolle mit fairen Verfahren

Wie begrenzen gemeinsame Standards und Einspruchsverfahren Missbrauch, ohne legitime Werke auszusperren?

Rechtsfälle zeigen den Handlungsbedarf

Streamingbetrug: Verurteilung

Michael Smith: 18 Monate Haft wegen KI-gestützten Streamingbetrugs (Oktober 2026).

Suno und Udio: Urheberrechtsstreit

Klagen wegen KI-Training mit geschützten Aufnahmen. Einige Verfahren endeten in Lizenzvereinbarungen.

DistroKid: Klage gegen den Distributor

UMG klagt wegen Täuschung und Urheberrechtsverletzungen. DistroKid bestreitet dies.

Alphabet: Klage gegen die Führung

Ein Aktionär klagt gegen Organmitglieder wegen behaupteter Risiken beim KI-Training.

Stand: 08.11.2026. Klagevorwürfe sind keine gerichtlichen Feststellungen.

© Fraunhofer IDMT

Contact



Hanna Lukashevich
Head of Semantic Media Technologies

+49 3677 467-224
hanna.lukashevich@idmt.Fraunhofer.de

Fraunhofer IDMT
Ehrenbergstrasse 31
98693 Ilmenau, Germany
www.idmt.Fraunhofer.de