

Fiche explicative – MusicLM

- ♦ **Nom du modèle :**

MusicLM

- ♦ **Type :**

Modèle de génération musicale **texte-à-musique** basé sur **Transformers** et **représentation audio latente**.

- ♦ **Développeur :**

Google Research (Brain Team + DeepMind)

- ♦ **Date de sortie :**

Publication du papier : **Janvier 2023**

Demo publique (test limité) : **Mai 2023**

Objectif

MusicLM vise à **générer de la musique complète à partir de simples descriptions textuelles** ("prompt textuel").

Contrairement à des modèles qui génèrent uniquement des notes MIDI ou du bruitage, **MusicLM génère directement des morceaux musicaux complexes** : mélodies, harmonies, instrumentation, ambiance sonore, et parfois même progression émotionnelle.

✅ Résultat : création de **musiques originales longues** (jusqu'à plusieurs minutes) **guidées par du texte**.

Fonctionnement simplifié

Étape	Description
Entrée	Texte descriptif ("jazz mélancolique au piano et basse douce")
Encodage	Le texte est converti en représentation latente audio
Génération	Modèle Transformer prédit la séquence latente correspondante
Reconstruction	Le modèle reconstruit l'audio final à partir du latent audio

Techniques utilisées :

- **Audio Tokenizer** : transforme l'audio en "tokens" compacts (SoundStream codec)
 - **MuLan** : encode le texte et l'audio dans un espace commun
 - **Hiérarchie de modèles** : un modèle grossier pour la structure, un modèle fin pour les détails
-

Applications concrètes

- **Génération musicale text-to-music** pour films, jeux vidéo, contenus numériques
 - **Aide à la composition** pour musiciens, producteurs
 - **Exploration créative** par la génération de styles inédits
-




Exemples d'usage

Domaine	Exemple
Production musicale	Générer des ambiances sonores ou des bases instrumentales originales
Jeux vidéo / VR	Créer des musiques adaptatives selon les scènes
Recherche en musique IA	Étudier comment l'IA comprend structure musicale et émotions

Détails techniques

Caractéristique	Valeur
Architecture	Transformer hiérarchique + encodage latent
Framework	TensorFlow + JAX (modèle interne Google)
Input	Prompt texte (langage naturel)
Output	Audio stéréo 24kHz, fichiers WAV
Dataset d'entraînement	5 millions d'heures de musique avec descriptions textuelles associées (MusicCaps, etc.)
Durée de génération	De quelques secondes à plusieurs minutes

Ressources officielles et utiles

-  [Publication scientifique officielle MusicLM \(arXiv\)](#)
 -  [Site officiel MusicLM Google Research](#)
 -  [Démonstration publique limitée sur MusicCaps Dataset](#)
-

Démonstrations & alternatives pratiques

Démonstration à tester (officielle)

-  [Exemples audio MusicLM \(Google\)](#)

Modèles alternatifs inspirés

-  [MusicGen \(Meta AI\) – inspiré de MusicLM et disponible publiquement !](#)

(Attention : MusicLM complet n'est pas open-source pour le public, mais plusieurs alternatives très proches comme MusicGen existent aujourd'hui.)



Tableau des avantages / inconvénients



Avantages

Génère des musiques longues (>1 minute)

Bonne cohérence harmonique et stylistique

Peut suivre des descriptions textuelles complexes

Exemples impressionnants publiquement accessibles



Inconvénients

MusicLM complet non disponible pour le public

Nécessite d'énormes ressources GPU/TPU

Contrôle précis sur la mélodie/dynamique limité

Possibles biais éthiques liés aux datasets musicaux