Módulo 5

GENERACIÓN Y EDICIÓN DE AUDIO

Contenido

INTRODUCCIÓN A LA GENERACIÓN DE AUDIO

INTRODUCCIÓN

Después de haber explorado en los módulos anteriores cómo la inteligencia artificial se aplica en el ámbito del texto y la imagen, ahora nos adentraremos en el mundo de la **generación y edición de audio asistida por IA**. La IA no solo automatiza tareas, sino que impulsa la creatividad en la producción de audio, facilitando la creación de sonidos personalizados o la edición de pistas complejas de manera rápida y eficiente. Este avance abre nuevas posibilidades para el desarrollo de contenido auditivo en áreas como la música, el entretenimiento y la publicidad.

IA en la generación de audio

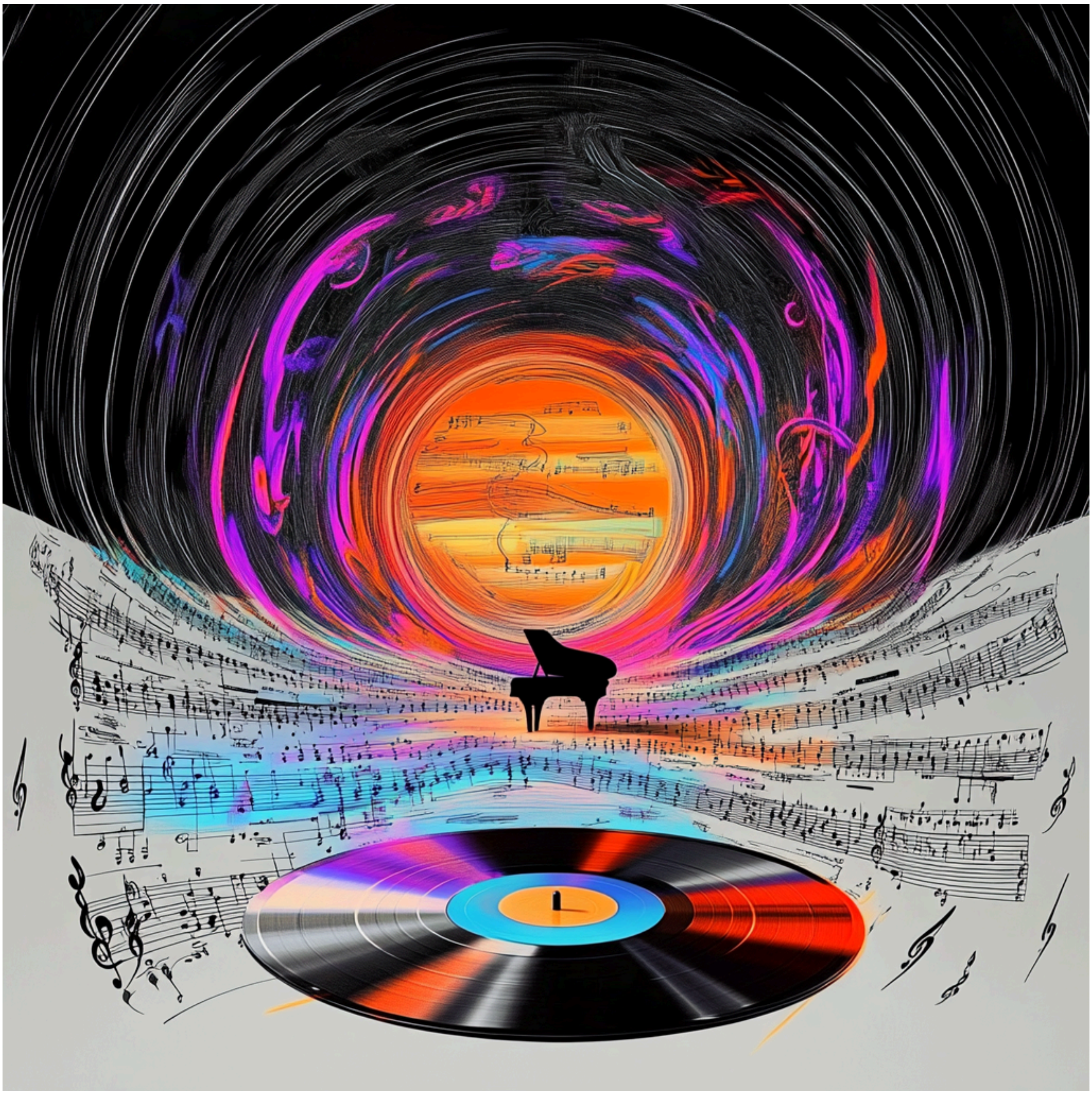
En el ámbito del audio, la IA está revolucionando la forma en que creamos y consumimos contenido sonoro. Los algoritmos de IA pueden analizar enormes bases de datos de audio y aprender a generar nuevos sonidos, ajustándose a diversos estilos. Por ejemplo, es posible crear una banda sonora personalizada para un videojuego o un anuncio, donde la IA genera la música y los efectos de sonido en función de unas pocas indicaciones, como el tono o el tipo de emoción que se desea transmitir.

Además de la creación de sonidos y melodías originales, la IA mejora el proceso de **remasterización de audio**. Esta tecnología puede restaurar grabaciones antiguas, eliminar ruidos no deseados y mejorar la calidad de archivos de sonido antiguos, ajustándolos a los estándares modernos sin perder su esencia original.

Tecnologías Fundamentales: Redes Neuronales y Redes Generativas

El mundo del audio está experimentando una transformación gracias a tecnologías como las **Redes Neuronales Artificiales** y los **Modelos Generativos** (como las redes GANs). Estas herramientas pueden generar contenido sonoro realista, incluso imitando las voces de personas o instrumentos musicales que no están realmente grabados. Un ejemplo de esto es la generación de voces sintéticas para asistentes virtuales o la creación de música en tiempo real para videojuegos.

Otra técnica popular es la **transferencia de estilo de audio**, que permite aplicar las características de un género musical o el tono de un locutor a una grabación nueva. Esto abre posibilidades creativas interesantes, como transformar una voz en un estilo particular o darle un toque específico a una melodía.



IA y creatividad en la producción de audio

La IA también facilita la personalización del contenido auditivo. Por ejemplo, una herramienta de IA puede ajustar una pista de música a las preferencias individuales del oyente o adaptar un podcast al estilo que prefiera su audiencia. Además, al automatizar tareas repetitivas como la edición de audio, la IA permite a los creadores centrarse en la verdadera innovación, aportando valor en los aspectos más artísticos del proceso creativo.

A lo largo de este módulo, exploraremos los fundamentos de estas tecnologías y aprenderemos a utilizar herramientas específicas que permiten generar y editar **audio asistido por IA**. Estas herramientas hacen que la creación de contenido auditivo sea accesible y emocionante, pero también es importante reflexionar sobre sus implicaciones éticas y cómo equilibrar la automatización con la creatividad humana.

Estamos viviendo un momento apasionante, donde la colaboración entre humanos y máquinas redefine los límites de la creatividad en el mundo del audio..

HISTORIA Y EVOLUCIÓN DE LA GENERACIÓN DE AUDIO POR IA

1. Los primeros pasos (1950s - 1980s)

Durante este periodo, la generación de audio comenzó con experimentos en procesamiento digital de señales y síntesis de audio. Los primeros sintetizadores analógicos y computadoras permitieron crear y manipular sonidos de forma básica, aunque aún no se consideraban IA en el sentido actual.

Por ejemplo, Max Mathews, en los años 50, desarrolló el software "*Music I*", el primer programa de computadora para sintetizar audio, abriendo el camino para futuras investigaciones en sonido digital.

2. El surgimiento de las Redes Neuronales (1980s - 1990s):

En los 80 y 90, los avances en redes neuronales permitieron a las máquinas analizar patrones en datos sonoros. Aunque limitadas, estas redes mejoraron la calidad de la generación de audio mediante técnicas de aprendizaje básico, permitiendo a las máquinas reconocer patrones en voz y música.

Durante estos años, surgieron los primeros modelos de reconocimiento de voz, que permitieron a las máquinas identificar palabras, aunque con precisión limitada.

3. Algoritmos de Análisis y Síntesis Avanzados (2000s):

Con el uso de redes neuronales profundas y modelos de sintetización de audio más avanzados, los algoritmos de IA pudieron identificar patrones en la música y el habla. Esto marcó el inicio de la IA en aplicaciones prácticas, como la conversión de texto a voz y la generación básica de música.

Entre otros avances, el sistema Auto-Tune, aunque más relacionado con la edición que con la generación, marcó un cambio en el procesamiento del audio al usar algoritmos para ajustar y modificar el tono de la voz en grabaciones de audio.

4. El auge de las Redes Generativas (GANs) y Modelos de Síntesis de Audio (2010s):

En los años 2010, con el desarrollo de redes como las Redes Generativas Adversarias (GANs), la IA comenzó a generar muestras de audio de forma más convincente. Las GANs permitieron a los modelos crear contenido sonoro que simulaba instrumentos y voces con mayor precisión.

En 2016, ***WaveNet*** de Google revolucionó la síntesis de voz con redes neuronales que generaban **voces naturales y expresivas**, permitiendo mejorar aplicaciones de texto a voz como los asistentes de voz.

5. Modelos Avanzados y Personalización (2015 - 2020):

Con el auge de los datos y la mejora en modelos generativos, surgieron herramientas para crear música y voces con un realismo sin precedentes. También aparecieron técnicas de transformación de estilo en audio, que aplican características de un estilo musical a otro.

OpenAI lanzó ***Jukebox***, un modelo capaz de generar canciones completas en diversos géneros, con voces y letras coherentes. También se perfeccionaron los modelos de *deepfake* para voz, usados para replicar voces de personajes históricos o celebridades.

6. Democratización del Audio Generado por Usuario (2020 y más allá):

Las herramientas de IA para la generación de audio se han vuelto más accesibles, permitiendo a cualquiera crear contenido sonoro desde casa. Aplicaciones y plataformas permiten generar música, voces y efectos de sonido con calidad profesional, democratizando el proceso.

Por ejemplo, herramientas como **Descript** y **Resemble.ai** han hecho accesible la clonación y síntesis de voz, mientras que plataformas de creación musical como **Amper Music** permiten a usuarios sin conocimientos musicales componer piezas únicas mediante IA.

De cara a los próximos años se prevé que la evolución en la generación de audio por IA continúe abriendo nuevas posibilidades especialmente en ámbitos como la música, cine, juegos y publicidad, permitiendo que creadores de todos los niveles puedan acceder a tecnologías de producción de audio de alta calidad.

APLICACIÓN DE LA IA EN DIFERENTES ÁMBITOS E INDUSTRIAS

La inteligencia artificial ha transformado la generación y edición de audio, permitiendo avances innovadores en distintos sectores. A continuación, exploramos cómo la IA está impactando en diversas industrias a través de la creación de contenido sonoro y musical::

1. Entretenimiento y Música:

La IA permite la creación de música original y personalizada, adaptándose a diferentes géneros y estilos. Modelos como **Jukebox** de OpenAI pueden generar canciones completas en distintos géneros, y herramientas como **Aiva** crean composiciones originales en estilos como el clásico o el jazz.

Por ejemplo, se están generando bandas sonoras personalizadas para películas y videojuegos, ajustando el tono y ritmo según las necesidades de cada escena, sin necesidad de compositores humanos..

2. Publicidad y marketing:

Las marcas utilizan la IA para generar pistas sonoras y jingles personalizados, adaptados a las preferencias y emociones de sus audiencias. La IA analiza datos de los usuarios para crear sonidos que capten su atención y mejoren la eficacia de las campañas publicitarias.

Un uso bastante habitual es la creación automática de música de fondo que refleja el tono de los anuncios en redes sociales, como música relajante para productos de bienestar o ritmos más dinámicos para promociones de productos tecnológicos.

3. Podcasts y Audiolibros:

La IA ha facilitado la generación de voces sintéticas y la edición automatizada de audios, haciendo posible producir audiolibros y podcasts con alta calidad en menor tiempo. Herramientas como **Descript** permiten realizar ediciones automáticas y mejorar la claridad de las grabaciones.

Es fácil encontrar ya herramientas capaces de crear voces narrativas para audiolibros y programas de podcast, que ajustan el tono y el estilo de la voz a la temática del contenido.

4. Atención al Cliente y Asistentes Virtuales:

En los servicios de atención al cliente, la IA genera voces personalizadas para asistentes virtuales y respuestas automáticas en aplicaciones de mensajería. Esto permite a las empresas ofrecer soporte rápido y eficiente en una variedad de idiomas y estilos de voz.

Entre diversos ámbitos, las empresas de telecomunicaciones y banca están usando IA para crear asistentes virtuales con voces amigables y profesionalmente entrenadas para resolver preguntas comunes de sus clientes.

5. Educación y Capacitación:

En la educación, la IA genera audios educativos como lecciones de idiomas, narraciones de textos educativos y explicaciones de conceptos complejos. Esto facilita el aprendizaje a través de contenido auditivo y adaptado a los distintos ritmos de estudio.

6. Medios y Periodismo:

La IA ayuda a producir resúmenes de noticias en formato de audio, adaptando la entonación y el ritmo a la seriedad o urgencia de las noticias. Los medios usan IA para crear boletines sonoros automáticos, facilitando la entrega de información actualizada.

También lo estamos encontrando en muchos blogs, que incluyen el clips de audio correspondiente al texto escrito.

7. Rehabilitación y Salud Mental:

La IA se utiliza en aplicaciones de **terapia de voz** y **rehabilitación auditiva**, creando audios personalizados para mejorar la dicción y la comprensión del lenguaje en personas con dificultades auditivas o del habla.

Por ejemplo se está empleando en la generación de sonidos relajantes para aplicaciones de meditación y terapia de sueño, como ruido blanco y música ambiental, ajustada para promover el bienestar emocional.

8. Industria de los Videojuegos:

La IA permite crear efectos de sonido adaptativos y bandas sonoras que responden en tiempo real a las acciones del jugador, mejorando la inmersión en el juego.

Generación de sonidos que cambian en función de la localización y el entorno del personaje en el juego, como el crujir de hojas en un bosque o el sonido de una multitud en una ciudad.

No es una lista exhaustiva, sino solo algunos de los ámbitos donde la generación de audio se está aplicando de manera más significativa. La IA no solo mejora la eficiencia en la producción de contenido de audio, sino que también permite una personalización inédita en los diferentes sectores, haciendo el sonido más accesible y adaptable a las necesidades de cada usuario y contexto.

CASOS DE ESTUDIO EN LA GENERACIÓN DE AUDIO ASISTIDO POR IA

La IA generativa está siendo utilizada en diversos sectores para transformar la producción de contenido auditivo. A continuación, algunos ejemplos de casos de estudio que muestran su impacto:

1. **Música Generativa: Amper Music en la Creación de Pistas Personalizadas**:

   Amper Music es una plataforma de IA que permite a los usuarios componer música personalizada sin conocimientos musicales previos. Usando algoritmos de aprendizaje automático, Amper genera música adaptada a diferentes géneros, tonos y duraciones, ajustándose a las preferencias del usuario.

   Esta herramienta se ha utilizado en la industria de los videojuegos y la publicidad, donde se necesita música adaptable y de bajo costo. Por ejemplo, creadores de contenido en redes sociales pueden generar pistas rápidas para sus vídeos, personalizando el sonido sin infringir derechos de autor.

2. **Audiolibros: DeepZen y la Narración Sintética**:

   DeepZen utiliza modelos de IA para generar voces que narran audiolibros de manera realista y con inflexiones emocionales. La IA puede replicar tonos y estilos de locución específicos, lo que permite la creación de audiolibros con voces personalizadas y expresivas.

   Este modelo ha permitido a editoriales reducir los costos y tiempos de producción en la creación de audiolibros, facilitando la disponibilidad de contenidos narrados en varios idiomas y estilos. Además, DeepZen ayuda a que autores autopublicados tengan acceso a narraciones de alta calidad sin contratar locutores.

3. **Atención al Cliente: Resemble AI en la Generación de Asistentes de Voz Personalizados**:

   Resemble AI permite a las empresas crear voces sintéticas que representan la marca en sus interacciones de atención al cliente. Estas voces se generan a partir de muestras reales y pueden expresar emociones y entonaciones específicas, mejorando la experiencia del cliente en interacciones telefónicas y asistentes virtuales.

   Empresas de telecomunicaciones y servicios financieros utilizan

Resemble AI para que sus asistentes de voz respondan de forma natural y personalizada, incrementando la satisfacción del cliente y reduciendo la necesidad de intervención humana en las consultas rutinarias.

4. **Terapia de Salud Mental: Endel y la Creación de Sonidos para Relajación**:

   La aplicación **Endel** utiliza IA para crear entornos sonoros que ayudan a reducir el estrés y mejorar el bienestar emocional. Basándose en datos del usuario, como el ritmo cardíaco y la hora del día, Endel genera paisajes sonoros personalizados que ayudan a la concentración, la relajación y el sueño.

   La app ha sido utilizada en entornos laborales y personales para mejorar la productividad y el bienestar. Es una de las primeras aplicaciones aprobadas científicamente para la terapia auditiva, y su uso se ha extendido en áreas de bienestar corporativo y terapias de salud mental.

Estos casos de estudio destacan cómo la IA en la generación de audio está facilitando el acceso a contenidos auditivos de alta calidad y permitiendo una personalización significativa en sectores como la música, los audiolibros, la atención al cliente y la salud mental.

Concierto compuesto por IA en la III edición ADDA

La Orquesta ADDA Sinfónica ha sorprendido a los asistentes del III Fórum Europeo de Inteligencia Artificial, con la interpretación de cuatro piezas creadas a través de IA y combinadas con imágenes también generadas por inteligencia artificial.

(Clic en imagen para ver vídeo:)

FUNDAMENTOS DE LA GENERACIÓN DE AUDIO ASISTIDO POR IA

En los últimos tiempos, hemos visto un notable crecimiento en herramientas y plataformas relacionadas con la generación de audio por IA, cambiando la forma en que creamos y experimentamos el sonido. Pero, ¿cómo puede una máquina aprender a generar audio? Vamos a explorar los principios fundamentales de estas tecnologías.

La **generación de vídeo por IA** se basa en una serie de técnicas de aprendizaje profundo, en las que las redes neuronales, como las **Redes Generativas Adversarias (GANs)** y los **Modelos de Difusión**, juegan un papel crucial. Estas redes aprenden patrones a partir de grandes conjuntos de datos, como secuencias de vídeo y películas, y luego son capaces de generar contenido visual nuevo y original que imita esos patrones.

Principios de la Generación de Audio por IA

La generación de audio por IA se basa en técnicas de aprendizaje profundo, en las que redes neuronales como las Redes Generativas Adversarias (GANs) y los Modelos de *Transformers* desempeñan un papel fundamental. Estos modelos aprenden a partir de grandes conjuntos de datos de audio, como grabaciones de voz, música y efectos sonoros. Al identificar patrones y estructuras en estos datos, la IA puede generar audio nuevo y original que imita las características de los sonidos en los que se ha entrenado.

Redes Generativas Adversarias (GANs):

Como ya hemos visto en módulos anteriores, las GANs consisten en dos redes: un generador que intenta crear audio realista y un discriminador que evalúa la calidad del audio generado. El proceso competitivo entre ambas redes permite que la IA aprenda a producir audio de alta fidelidad, capaz de replicar voces o instrumentos con un nivel de realismo sorprendente.

Por ejemplo, se usan en la creación de voces sintéticas para asistentes virtuales o doblajes automatizados en diferentes idiomas.

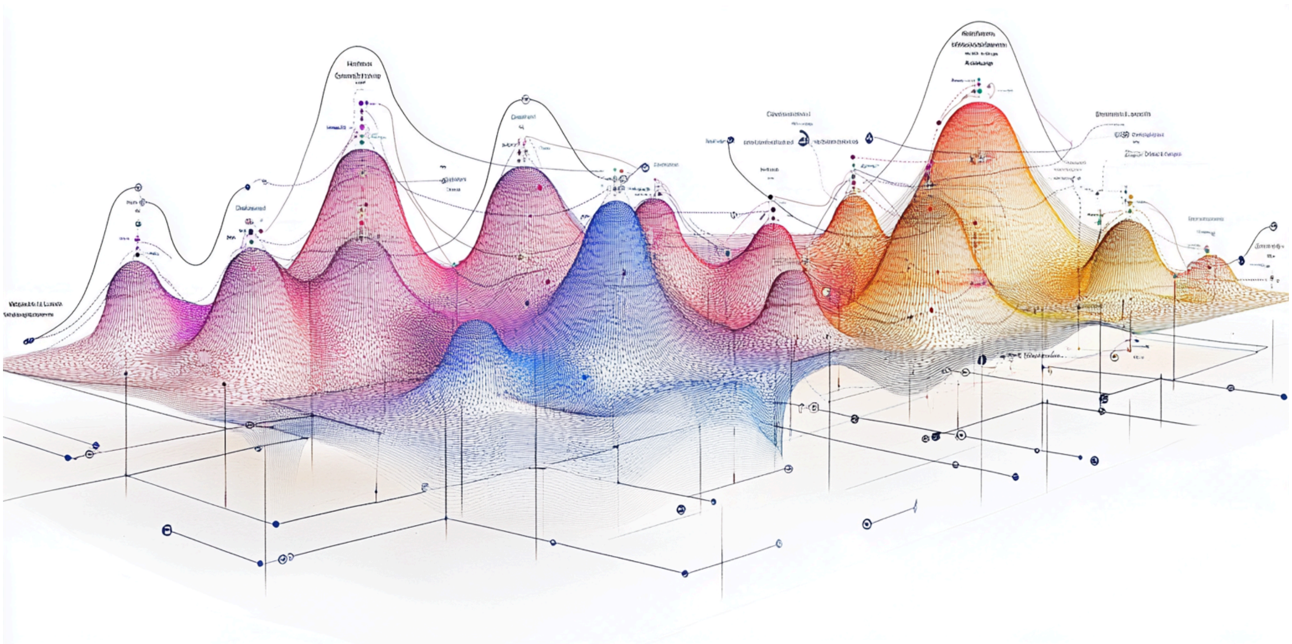
Modelos de *Transformers*:

Los *transformers* han demostrado ser especialmente efectivos en la generación de secuencias de audio coherentes, manteniendo el flujo y la estructura temporal. Este tipo de red neuronal es capaz de generar texto hablado o música, permitiendo a la IA mantener la coherencia en fragmentos de audio más largos.

Son conocidas herramientas como Jukebox de OpenAI, que generan canciones completas basadas en indicaciones textuales y mantienen una estructura armónica.

Modelos de Difusión:

En audio, los modelos de difusión generan el sonido en etapas, mejorando la precisión y eliminando gradualmente el ruido hasta que se obtiene un sonido claro. Esto es útil para crear música o efectos sonoros, especialmente en ambientes virtuales donde la fidelidad es esencial.



Entrenamiento y Aprendizaje de Patrones

Los modelos generativos de audio necesitan analizar grandes cantidades de datos para aprender patrones específicos de diferentes tipos de sonido. A través del entrenamiento, la IA capta características como el tono, la duración, el timbre y la dinámica. Esto le permite generar no solo voces y música, sino también efectos complejos que antes sólo podían producirse manualmente.

A continuación, los pasos clave en el aprendizaje de una IA para la generación de audio:

1. Entrenamiento con Datos Masivos

La IA se entrena con grandes colecciones de datos de audio, que incluyen muestras de voz, música y efectos sonoros. Estos datos ayudan a la IA a identificar características acústicas como el tono, el ritmo y el timbre.

Por ejemplo, para generar voces realistas, la IA analiza miles de grabaciones de diferentes personas, idiomas y emociones. Esto le permite capturar variaciones en la pronunciación y en las inflexiones de voz.

2. Aprendizaje de Patrones de Sonido

Al analizar estos datos, la IA aprende patrones específicos, como los cambios en el tono de una conversación, los ritmos en distintos géneros musicales o los efectos de reverberación en diferentes entornos. A través de este proceso, la IA adquiere la capacidad de imitar estilos de música y voces con coherencia.

En un modelo de generación de música, la IA aprende los patrones característicos de un género específico, como el jazz o el rock, lo que le permite generar canciones que suenan como piezas reales de ese género.

3. Generación de Contenido Nuevo

Con la información adquirida, la IA puede generar audio nuevo y original, ya sea una voz que suena como una persona en particular, una canción en un estilo musical específico o un sonido ambiental para un entorno virtual. Al recibir una descripción textual o un "prompt" que indique el tipo de audio deseado, la IA puede combinar las características aprendidas para crear contenido sonoro que sigue el estilo y la estructura de los datos originales.

Por ejemplo, al introducir un prompt como "canción de jazz relajante", la IA puede generar una pieza de jazz con patrones rítmicos y melódicos típicos de este género, manteniendo coherencia en la estructura de la canción.

Este enfoque de aprendizaje permite a las IA en audio crear contenido original y personalizado que puede usarse en música, efectos de sonido, voces sintéticas y más, ampliando las posibilidades en la industria del sonido y la producción musical.

Principios clave de la generación de audio por IA:

Los modelos generativos son fundamentales para la creación de contenido de audio. Estos programas no siguen simplemente reglas rígidas, sino que aprenden de datos reales para identificar patrones y relaciones, generando así sonidos que replican las características de los datos de entrenamiento. Esto permite a la IA producir audio nuevo que mantiene la esencia de los datos originales en términos de tono, ritmo y estructura sonora.

Por ejemplo, si entrenas un modelo generativo con una colección de música jazz, la IA aprenderá patrones específicos como el ritmo sincopado, la variación en los acordes y la improvisación característica del jazz. Así, la IA podría generar nuevas piezas de jazz que mantengan la coherencia y el estilo propios de este género.

Limitaciones y desafíos en la generación de audio por IA

La generación de audio por IA ofrece nuevas oportunidades en diversos sectores, pero también plantea varios desafíos técnicos, éticos y legales. A continuación, se analizan las principales limitaciones que enfrenta esta tecnología emergente:

1. Calidad y Realismo

Aunque los avances en la generación de audio por IA han mejorado la calidad, aún persisten problemas de realismo en **voces** y **música**. Algunas voces sintéticas pueden sonar robóticas, especialmente en frases complejas, y es difícil lograr las inflexiones y tonos naturales del habla humana en tiempo real.

2. Consumo de Recursos

La generación de audio por IA requiere un procesamiento computacional significativo, lo que implica un consumo elevado de energía y recursos para entrenar y ejecutar modelos avanzados. Esto puede ser costoso y tiene un impacto ambiental.

Entrenar modelos como WaveNet o Jukebox, de OpenAI, está exigiendo infraestructuras potentes y tiempos de procesamiento prolongados.

3. Sesgo en los datos

Si los datos de entrenamiento contienen sesgos culturales o de lenguaje, la IA replicará esos patrones, lo que puede resultar en audio generado que no refleje la diversidad del habla humana. Las voces pueden, por ejemplo, sonar más naturales en ciertos acentos y menos en otros.

Por ejemplo las herramientas de síntesis de voz que se entrenan solo con muestras de inglés estándar pueden generar menos precisión al intentar emular acentos o idiomas diferentes.

4. Derechos de Autor

La propiedad de contenido generado por IA es un área legal aún sin resolver. Si una canción o una voz sintética es creada por IA, surgen preguntas sobre quién posee los derechos de ese contenido: ¿el usuario, el desarrollador del modelo o el propietario de los datos de entrenamiento?

Actualmente la música generada por IA, como en el caso de Jukebox, plantea dudas sobre la propiedad, especialmente cuando el modelo emula estilos de artistas conocidos.

5. Ética y uso malintencionado

Las herramientas de clonación de voz o *deepfakes* de audio pueden ser utilizadas de manera malintencionada, como en fraudes telefónicos o en la generación de declaraciones falsas. Esto plantea serios problemas éticos y de seguridad.

Ya estamos comprobando cómo se están empleando voces generadas que imitan a figuras públicas que son usadas para difundir información errónea o manipular la opinión pública

6. Interpretabilidad y Control

Las redes neuronales profundas y los modelos generativos son difíciles de interpretar. Esto significa que, en caso de errores o resultados inesperados, los desarrolladores pueden no entender completamente cómo se generó un determinado sonido o música, limitando el control y la precisión.

Por ejemplo, en un proyecto de doblaje automático, un modelo de IA podría generar interpretaciones de voz poco precisas o inadecuadas, sin una explicación clara de las causas.

7. Calidad de los datos de entrenamiento

La IA depende de grandes cantidades de datos de audio de alta calidad para entrenarse de forma efectiva. Si los datos de entrenamiento son limitados o no representan una amplia gama de tonos y estilos, el audio generado también será limitado y carecerá de diversidad.

Algunas voces generadas pueden sonar planas o poco expresivas cuando los datos de entrenamiento no incluyen variaciones emocionales o tonos diferentes.

8. Accesibilidad y Costos

Las tecnologías de generación de audio por IA avanzadas pueden ser costosas y no están siempre disponibles para pequeñas empresas o creadores individuales. Esto limita la democratización de la tecnología y su uso generalizado.

De hecho el acceso a herramientas como Jukebox o Resemble AI a nivel de empresa puede ser prohibitivo para algunos usuarios, lo que reduce la posibilidad de innovar en proyectos de menor escala.

CRÉDITOS

El contenido de este módulo ha sido elaborado por Fran Bartolomé V-Gamazo, profesor y consultor homologado de la Escuela de Organización Industrial, especialista en Robótica, Inteligencia Artificial, Programación e Impresión 3D.

Las imágenes han sido creadas por IA en Midjourney, salvo las relacionadas con el concierto compuesto por IA en la III edición ADDA.