

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



**MODELADO Y GENERACIÓN  
MUSICAL: UNA APROXIMACIÓN  
MEDIANTE REDES NEURONALES Y  
ENTRADAS MIDI**

TESINA

QUE PARA OBTENER EL TÍTULO DE

INGENIERÍA EN MECATRÓNICA

PRESENTA

SANTIAGO PÉREZ GUTIÉRREZ

ASESOR

WILMER EFRÉN PEREIRA GONZÁLEZ

# Tabla de contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Contexto . . . . .	1
1.2. Identificación del problema . . . . .	2
1.3. Objetivo . . . . .	3
1.4. Alcance . . . . .	3
1.5. Metodología . . . . .	7
1.5.1. Detectar brecha: . . . . .	9
1.5.2. Definir el problema: . . . . .	9
1.5.3. Explorar alternativas: . . . . .	9
1.5.4. Seleccionar plan: . . . . .	9
1.5.5. Documentación y Diseminación: . . . . .	10
1.6. Organización del documento . . . . .	10
<b>2. Análisis</b>	<b>12</b>
2.1. Requerimientos Funcionales . . . . .	12
2.2. Requerimientos No Funcionales . . . . .	14
2.3. Restricciones Funcionales . . . . .	15
2.4. Restricciones de Datos: . . . . .	17
2.5. Trabajos relacionados . . . . .	18
2.5.1. Revisión de la Literatura: . . . . .	18

<i>TABLA DE CONTENIDO</i>	2
2.5.2. Comparación de Modelos: . . . . .	19
<b>3. Diseño</b>	<b>21</b>
3.1. Arquitectura . . . . .	21
3.1.1. Detalles de la Red Neuronal . . . . .	21
3.1.2. Manejo de Entradas y Salidas . . . . .	22
3.1.3. Regularización y Prevención del Sobreajuste . . . . .	24
3.1.4. Flujo de Datos y Procesamiento . . . . .	25
3.2. Soluciones Alternativas . . . . .	25
<b>4. Implementación</b>	<b>26</b>
4.1. Modelo óptimo . . . . .	26
<b>5. Pruebas y resultados</b>	<b>27</b>
5.1. Pruebas realizadas y sus resultados . . . . .	27
<b>6. Conclusiones</b>	<b>29</b>
6.1. Conclusiones pruebas de velocidad y posición . . . . .	29
<b>Referencias</b>	<b>31</b>

# Capítulo 1

## Introducción

### 1.1. Contexto

La música ha sido un reflejo de la evolución cultural y tecnológica de la humanidad desde tiempos inmemoriales. Desde las melodías simples creadas con instrumentos rudimentarios en culturas antiguas hasta las complejas composiciones orquestales del Renacimiento y más allá. La música ha servido como medio de expresión, comunicación y unión para las sociedades de todo el mundo.

Con la Revolución Industrial y los avances tecnológicos del siglo XX, la música experimentó una metamorfosis significativa. Surgieron nuevos instrumentos, métodos de grabación y formas de distribución, transformando radicalmente la industria musical. Una de estas revoluciones tecnológicas fue la introducción del formato MIDI en la década de 1980. Esta interfaz digital permitió una representación precisa de la música en términos de notas, duraciones, intensidades y otros parámetros. El MIDI no solo democratizó la producción musical, permitiendo a músicos sin acceso a grandes estudios producir música de alta calidad, sino que también sentó las

bases para la experimentación y manipulación digital de la música.

## 1.2. Identificación del problema

En el amanecer del siglo XXI, la integración de la inteligencia artificial en el ámbito musical ha abierto nuevos horizontes de exploración y creatividad. A medida que la tecnología avanza, el desafío de fusionar la complejidad de la expresión musical con la precisión de la inteligencia artificial se ha vuelto cada vez más prominente. En particular, el uso de redes neuronales para interpretar y generar música a partir de entradas MIDI presenta un conjunto único de desafíos y oportunidades. Este problema se centra en cómo podemos diseñar redes neuronales capaces de comprender la profundidad y la variabilidad de la música, transformando secuencias de datos MIDI en composiciones que no solo emulan la estructura musical humana, sino que también capturan la esencia de la creatividad y la emoción.

La tarea específica enfrenta obstáculos tanto técnicos como creativos. Técnicamente, la red debe ser capaz de procesar la información MIDI, que codifica aspectos como la nota, la duración, la intensidad y el timbre, y utilizar esta información para generar música que sea coherente, armónicamente rica y estilísticamente variada. Creativamente, el desafío radica en ir más allá de la simple emulación de patrones musicales existentes para generar obras que sean genuinamente innovadoras y expresivas.

Por lo tanto, el problema central es doble: ¿Cómo podemos desarrollar algoritmos de inteligencia artificial que no solo comprendan la complejidad teórica y estructural de la música a través de entradas MIDI, sino que también incorporen elementos de creatividad y expresividad que caracterizan a la música creada por humanos? Este reto no solo busca cerrar la brecha entre la capacidad computacional y la creatividad humana sino

también explorar nuevas dimensiones de la creación musical, ampliando nuestras concepciones de lo que la música puede ser

### 1.3. Objetivo

El objetivo de este proyecto es desarrollar una red neuronal que pueda analizar y expandir fragmentos musicales en formato MIDI. Queremos que esta red no solo genere nuevas piezas musicales basándose en lo que ha aprendido, sino que también asegure que estas creaciones sean coherentes con el estilo y estructura del fragmento original. La intención es crear música que no solo suene bien, sino que también respete y refleje el estilo del material de entrada. En esencia, buscamos combinar la capacidad técnica de la inteligencia artificial con un toque de creatividad y sensibilidad musical, para producir composiciones que sean tanto innovadoras como fieles a la esencia de la pieza original.

### 1.4. Alcance

Este estudio abarcará varias áreas interdisciplinarias, incluyendo teoría musical, procesamiento digital de señales y técnicas avanzadas de aprendizaje automático. Los principales aspectos cubiertos incluyen:

- **Investigación Teórica:** Se llevará a cabo una revisión exhaustiva de la literatura sobre generación musical asistida por IA, con un énfasis en trabajos anteriores que utilizan formatos MIDI y redes neuronales. Esta revisión ayudará a contextualizar el trabajo actual dentro del panorama más amplio del campo y a identificar áreas de oportunidad.
- **Desarrollo Técnico:** Este segmento abordará una comparativa exhaustiva de las diversas arquitecturas de redes neuronales

disponibles, con el objetivo de identificar la opción más adecuada para la construcción de nuestro modelo. Posteriormente, procederemos al diseño y desarrollo de una red neuronal meticulosamente adaptada para analizar y sintetizar música a partir de datos MIDI. Este proceso involucrará la selección cuidadosa de la arquitectura óptima, junto con estrategias avanzadas para el preprocesamiento de datos y técnicas de entrenamiento innovadoras, garantizando así la eficacia y eficiencia del modelo en la generación de música coherente y estilísticamente congruente.

- **Comparación con Modelos Existentes:** Para destacar las capacidades únicas de nuestro modelo, realizaremos un análisis comparativo detallado con modelos preentrenados reconocidos y accesibles a todo público, como lo son modelos en plataformas como HuggingFace o Kaggle. Este análisis se centrará en evaluar la calidad musical y la coherencia de las composiciones generadas, sirviendo como una medida objetiva para entender las mejoras y contribuciones significativas de nuestro modelo. Este proceso no solo subraya el valor agregado de nuestro trabajo sino también fortalece la confianza en las innovaciones propuestas.
- **Evaluación:** Tras la culminación del desarrollo de nuestra red neuronal, emprenderemos un exhaustivo proceso de evaluación para medir meticulosamente la calidad de la música generada. Este proceso abarcará tanto análisis cuantitativos como cualitativos, asegurando una valoración completa del modelo.

**Métricas Objetivas:** Nos apoyaremos en métricas objetivas para evaluar aspectos técnicos de la producción musical, tales como la coherencia melódica y armónica. Esto implica utilizar indicadores precisos para cuantificar la habilidad del modelo en replicar o innovar

dentro de los patrones musicales reconocibles, garantizando que la música generada cumpla con estándares de calidad y complejidad musical. Así mismo, evaluaremos que las notas generadas por el modelo, respeten los acordes y escala musical que el usuario genere con el instrumento MIDI.

**Evaluaciones Subjetivas por Expertos:** Complementariamente, invitaremos a expertos en música para que realicen evaluaciones subjetivas de las composiciones. Este enfoque nos permitirá captar la apreciación estética y emocional de la música generada, ofreciendo perspectivas valiosas sobre la resonancia y el impacto de la música en oídos experimentados.

Este marco de evaluación multidimensional nos permitirá no solo validar la eficacia técnica del modelo sino también su capacidad para innovar y emocionar, asegurando que las composiciones generadas no solo sean técnicamente competentes, sino también musicalmente enriquecedoras.

- **Aplicaciones Potenciales:** Anticipamos una gama de aplicaciones potenciales que podrían revolucionar la forma en que interactuamos con la música. Este modelo promete ser una herramienta invaluable en diversos ámbitos, desde la asistencia en la composición musical hasta su uso en educación y entretenimiento.

En el campo de la composición, el modelo puede llegar a ser un aliado para músicos y compositores, proporcionando soporte creativo mediante la sugerencia de melodías, armonías y ritmos. Esta asistencia no solo agilizará el proceso creativo sino que también abrirá nuevas avenidas para la experimentación musical.

Dentro del ámbito educativo, el modelo se proyecta como una herramienta pedagógica de vanguardia. Al permitir a los estudiantes



interactuar con un sistema capaz de generar música en una variedad de estilos a partir de unas pocas notas, esperamos fomentar una comprensión más profunda de la teoría musical y estimular la creatividad estudiantil, facilitando un aprendizaje interactivo y dinámico.

En lo que respecta al entretenimiento, las posibilidades son igualmente emocionantes. Podría utilizarse para crear bandas sonoras personalizadas para videojuegos, películas y otras aplicaciones multimedia, o incluso generar música ambiental personalizada para eventos y espacios públicos, adaptándose a las preferencias específicas del público o ambiente.

- Limitaciones: Este estudio se propone explorar un vasto espectro de géneros y estilos musicales, evidenciando la flexibilidad y capacidad de adaptación de nuestro modelo ante diferentes contextos musicales. A pesar de nuestros esfuerzos por abarcar una amplia gama de expresiones musicales, debemos reconocer que la diversidad musical es tan extensa que limita nuestra capacidad para incluir todos los géneros y contextos. Además, enfrentamos restricciones técnicas al trabajar con entradas MIDI, las cuales, por su naturaleza, no capturan la totalidad de los matices que caracterizan a las composiciones musicales más complejas.

Un aspecto fundamental que no debe pasarse por alto son las implicaciones en materia de derechos de autor y los desafíos legales potenciales al utilizar melodías, armonías o ritmos específicos dentro de nuestro modelo. La generación de contenido musical mediante inteligencia artificial plantea preguntas importantes sobre la originalidad y la propiedad intelectual, especialmente cuando se replican estilos o se inspiran en obras protegidas por derechos de

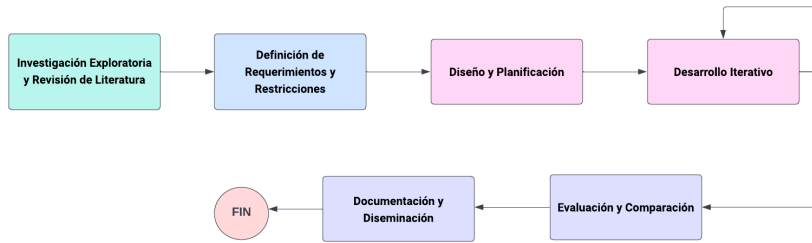
autor. Este estudio incluye un análisis crítico de las limitaciones técnicas, específicamente en lo que respecta a las redes neuronales y el formato MIDI, al tiempo que aborda las preocupaciones legales relacionadas con la generación musical automatizada.

Nuestro enfoque intenta no solo demostrar las capacidades del modelo sino también garantizar una discusión honesta y transparente sobre sus restricciones, incluyendo la navegación por el complejo panorama de los derechos de autor. Al considerar estas cuestiones éticas y legales, reafirmamos nuestro compromiso con la responsabilidad y la integridad en la innovación tecnológica.

## 1.5. Metodología

- Investigación Exploratoria y Revisión de Literatura: Revisar literatura académica y artículos de la industria. Estudiar modelos existentes y su funcionamiento. Identificar brechas y oportunidades para la innovación.
- Definición de Requerimientos y Restricciones: Definir los requerimientos funcionales y no funcionales. Identificar las restricciones técnicas y de datos.
- Diseño y Planificación: Seleccionar las arquitecturas de red neuronal y técnicas de procesamiento de datos MIDI adecuadas. Planificar las fases de desarrollo, pruebas y evaluación.
- Desarrollo Iterativo: Realizar pruebas cuantitativas y cualitativas. Comparar con modelos preentrenados y estudios relacionados. Recopilar feedback de usuarios y expertos.

- Documentación y Diseminación: Redactar informes detallados y documentos de investigación. Preparar presentaciones para compartir los hallazgos.



**Figura 1.4.1:** Metodología de Trabajo

### **1.5.1. Detectar brecha:**

Se realizará una revisión sistemática de la literatura para identificar las lagunas en la investigación actual relacionadas con la generación de música utilizando IA y MIDI. Esta revisión no solo se centrará en lo que se ha hecho, sino también en lo que falta en el campo, tales como limitaciones en estilos musicales, diversidad de datos y rendimiento de los modelos existentes.

### **1.5.2. Definir el problema:**

Con base en la revisión de la literatura, se establecerá una definición clara del problema que incluirá los desafíos específicos a abordar con la investigación. Esto incluirá la determinación de los requisitos técnicos y musicales para el modelo de red neuronal y la definición de qué constituye una generación musical exitosa

### **1.5.3. Explorar alternativas:**

Investigar una variedad de arquitecturas de red neuronal, como las redes neuronales convolucionales (CNN), redes recurrentes (RNN), y las redes generativas adversarias (GAN), para determinar cuál es más adecuada para el modelado de secuencias musicales MIDI. Además, se evaluarán diferentes métodos de preprocesamiento de datos MIDI y se determinará cómo influirán en la calidad de la salida musical.

### **1.5.4. Seleccionar plan:**

Seleccionar la arquitectura de red neuronal más prometedora y desarrollar un plan para su implementación. Esto incluirá:

- Selección de Datos: Identificar y curar un conjunto de datos de entrenamiento de alta calidad que sea representativo de los estilos

musicales que el modelo pretende generar.

- **Desarrollo y Entrenamiento:** Implementar la red neuronal seleccionada y entrenarla con el conjunto de datos. Esto también incluirá la optimización de hiperparámetros y la evaluación de la red durante el entrenamiento para asegurar la convergencia.
- **Validación y Pruebas:** Validar el modelo contra un conjunto de datos de prueba que no se haya utilizado durante el entrenamiento. Esto incluirá pruebas de robustez y la capacidad del modelo para generalizar a partir de nuevas entradas MIDI.
- **Comparación de Modelos:** Ejecutar una comparación detallada con modelos preentrenados existentes para evaluar el rendimiento relativo. Se utilizarán métricas como la precisión melódica, armónica y rítmica, así como la originalidad y la calidad estética percibida.
- **Evaluación del Usuario:** Realizar evaluaciones cualitativas con músicos y compositores para obtener retroalimentación sobre la utilidad y la musicalidad de las composiciones generadas.

#### **1.5.5. Documentación y Disseminación:**

Documentar todos los procesos y resultados de manera detallada para permitir la reproducibilidad y proporcionar una base sólida para futuras investigaciones.

### **1.6. Organización del documento**

El documento está estructurado de la siguiente forma:

- **Introducción:** Descripción de la situación, contexto, causas y metodología a seguir para llegar a la solución y concluir con el objetivo de la tesina.
- **Análisis del problema:** Características generales que debe cumplir la tesina, así como las restricciones que presenta para el desarrollo, algunos trabajos que realizan una función similar y las principales diferencias que presenta la tesina con respecto a los trabajos relacionados.
- **Definición del sistema a implementar:** Detalles de la arquitectura bajo la cual funciona el sistema desarrollado así como diferentes formas de implementar la solución y los distintos estándares utilizados en el sistema.
- **Propuesta de solución al sistema:** Descripción de los principales componentes, la implementación y diseño del sistema, así como las especificaciones de software y hardware utilizado para cumplir con el objetivo de la tesina.
- **Pruebas y resultados:** Pruebas acerca de viabilidad del proyecto, las configuraciones realizadas y el cumplimiento de los requerimientos planteados al inicio.
- **Conclusiones:** Conclusiones obtenidas del trabajo y los resultados obtenidos, así como observaciones a trabajos futuros que se pueden realizar y optimización del proyecto.

## Capítulo 2

# Análisis

En esta sección, delineamos los criterios esenciales que nuestro sistema de generación musical impulsado por inteligencia artificial debe satisfacer para ser considerado exitoso. Estos requisitos fundamentales servirán como pilares para el diseño, desarrollo e implementación del modelo, estableciendo un marco claro para medir su rendimiento y eficacia.

### 2.1. Requerimientos Funcionales

Para asegurar que nuestro modelo de inteligencia artificial alcance su máximo potencial en la generación de música, hemos establecido una serie de requerimientos funcionales clave. Estos requisitos orientarán el proceso de desarrollo, permitiendo que el modelo produzca resultados musicales de alta calidad y relevantes para una amplia gama de aplicaciones.

- **Precisión Melódica:** El modelo estará diseñado para generar secuencias melódicas que reflejen fielmente el espíritu y la estructura del fragmento MIDI inicial. Esto incluye adherirse a la tonalidad original, replicar el contorno melódico y seguir las tendencias

estilísticas de la música de origen, asegurando que cada pieza generada sea una extensión coherente y armoniosa del material inicial.

- **Armonización Adecuada:** Esperamos que el modelo produzca secuencias armónicas que no solo complementen las líneas melódicas sino que también se alineen con los acordes y progresiones característicos del género musical específico del fragmento de origen. La capacidad para generar armonías que respeten y enriquezcan la pieza original es esencial para la autenticidad y disfrute de la música creada.
- **Consistencia Rítmica:** Será crucial que el modelo mantenga una coherencia rítmica a lo largo de la composición, replicando los patrones rítmicos del fragmento inicial y expandiéndolos de manera que la música generada ofrezca un flujo natural y reconocible. Esto contribuirá a una experiencia auditiva fluida y agradable.
- **Generación Continuada:** El sistema deberá ser capaz de generar composiciones musicales continuas a partir de un fragmento MIDI inicial, creando transiciones suaves y lógicas que mantengan la integridad estilística y temática sin interrupciones abruptas. La continuidad es fundamental para lograr piezas musicales completas y cohesivas.
- **Versatilidad Estilística:** Nuestro modelo demostrará una notable capacidad para adaptarse a una diversidad de estilos musicales, desde jazz hasta clásico, pop y más allá. Deberá identificar y aplicar las características distintivas de cada género para generar música que sea auténticamente representativa de su estilo, permitiendo una generación flexible y personalizada según las preferencias del usuario.



Al cumplir con estos requerimientos funcionales, nuestro modelo no solo será capaz de generar música que sea técnicamente competente y estilísticamente diversa, sino que también proporcionará una herramienta poderosa y versátil para la creación musical, enriqueciendo el campo de la música generada por IA.

## 2.2. Requerimientos No Funcionales

Para asegurar que nuestro modelo de generación musical cumpla con las expectativas de rendimiento y adaptabilidad en diversos contextos de uso, es esencial establecer claros requerimientos no funcionales. Estos criterios son fundamentales para el diseño y la implementación exitosa del sistema.

- **Tiempo de Respuesta:** Es primordial que nuestro modelo sea capaz de generar música dentro de un marco temporal razonable. Para aplicaciones que demandan interactividad en tiempo real, como actuaciones en vivo o experiencias de usuario interactivas, el modelo debe ofrecer respuestas ágiles, con latencias mínimas que no superen los límites perceptibles por los usuarios. Esto garantizará una experiencia fluida y sin interrupciones.
- **Eficiencia Computacional:** Nuestra solución debe optimizar el uso de recursos computacionales para asegurar una generación musical de alta calidad sin la necesidad de recurrir a hardware excesivamente especializado. Esto implica un diseño inteligente que equilibre complejidad y rendimiento, permitiendo así la ejecución eficiente tanto en sistemas de gama alta como en configuraciones más modestas.
- **Capacidad de Entrenamiento:** Nuestro sistema debe ser capaz de adaptarse y aprender efectivamente, incluso frente a limitaciones en

el volumen de datos disponibles para entrenamiento. Para superar este desafío, se incorporarán estrategias avanzadas como el aprendizaje por transferencia y el aumento de datos, las cuales están diseñadas para potenciar el aprendizaje del modelo y su capacidad para generalizar a partir de conjuntos de datos más pequeños o menos variados. Estas técnicas no solo mejoran la eficacia del entrenamiento sino que también amplían las posibilidades de aplicación del modelo en una variedad de contextos musicales.

Estos requerimientos no funcionales son fundamentales para el diseño y desarrollo de un modelo en generación musical que sea práctico, accesible y versátil. Al adherirnos a estos estándares, nos aseguramos de que nuestro proyecto no solo sea innovador desde el punto de vista técnico, sino también aplicable y valioso en situaciones del mundo real.

## 2.3. Restricciones Funcionales

Nuestro proyecto enfrenta ciertas restricciones que podrían influir en la amplitud del estudio, la selección de metodologías y los resultados obtenidos. Estas restricciones se dividen en dos categorías principales: técnicas y relacionadas con los datos.

- Capacidad de Cómputo Disponible: La eficacia y eficiencia de nuestro modelo de generación musical se ve significativamente afectada por la capacidad de procesamiento a nuestra disposición. La disponibilidad limitada de recursos computacionales puede restringir no solo la complejidad del modelo propuesto sino también la velocidad de entrenamiento y la habilidad para procesar datos en tiempo real. Estrategias como la optimización de modelos, el uso de hardware especializado o la computación en la nube podrían mitigar

estos efectos. Asimismo, el objetivo del modelo es poder eficientar los recursos computacionales tanto para el entranamiento como para su uso.

- **Limitaciones del Formato MIDI:** A pesar de ser un estándar en la música digital, el formato MIDI tiene sus limitaciones, especialmente en lo que respecta a capturar la riqueza de expresión musical como el timbre y la dinámica. Estas restricciones pueden afectar la profundidad y expresividad de la música generada por nuestro modelo. La exploración de técnicas avanzadas de síntesis sonora y modelado de expresiones podría ofrecer soluciones parciales a estas limitaciones.
- **Duración de Entrenamiento del Modelo:** El proceso de entrenamiento para modelos de inteligencia artificial, particularmente aquellos basados en redes neuronales profundas, puede ser extenso. Las limitaciones en este aspecto pueden influir directamente en la selección de arquitecturas de modelos más eficientes y en la cantidad de datos de entrenamiento empleados. La implementación de técnicas de entrenamiento acelerado o el ajuste fino de modelos preentrenados son posibles vías para superar estos desafíos.
- **Escalabilidad:** Un aspecto crítico de nuestro modelo es su capacidad para adaptarse y manejar aumentos tanto en el volumen del conjunto de datos como en la complejidad de las tareas de generación musical, sin sufrir una degradación significativa en el rendimiento. La elección de arquitecturas de modelos escalables y el diseño de sistemas que permitan una fácil expansión serán fundamentales para asegurar la viabilidad a largo plazo del proyecto. Una vez terminado el modelo, nos enfrentamos al problema de la escalabilidad de este y como

podemos hacer llegar este modelo a la mayor cantidad de músicos posibles.

## 2.4. Restricciones de Datos:

- Disponibilidad y Diversidad de los Datos MIDI: Un factor clave para la efectividad de nuestro modelo de generación musical reside en la riqueza y variedad del conjunto de datos de entrenamiento. La limitada disponibilidad de datos MIDI de acceso abierto y diversificados puede presentar desafíos significativos para la capacidad de entrenamiento de nuestro modelo y por lo tanto, de generalizar y crear música en una amplia gama de estilos. Es crucial para nosotros explorar y adquirir conjuntos de datos que abarquen un espectro amplio de géneros musicales, asegurando así que el modelo pueda aprender y reproducir una diversidad de expresiones musicales o por otro lado, centrar nuestro enfoque en solo un género musical.
- Derechos de Autor y Uso de Material Protegido: Dado que la música está protegida por derechos de autor, el uso de composiciones con restricciones legales para entrenar o probar nuestro modelo podría conllevar complicaciones jurídicas. Es imperativo garantizar que todos los datos MIDI empleados respeten las normativas de propiedad intelectual, optando por material con licencias adecuadas o de dominio público. Esta precaución no solo es legalmente necesaria, sino que también respalda las prácticas éticas en la investigación y desarrollo de tecnologías de IA.
- Sesgo de Datos: La selección de los conjuntos de datos para entrenar nuestro modelo puede introducir sesgos no deseados, limitando potencialmente la habilidad del modelo para generar ciertos tipos de

música o para reflejar adecuadamente una variedad de estilos y contextos culturales. Es esencial adoptar un enfoque crítico en la recopilación de datos, esforzándose por incluir una representación equitativa de géneros musicales y tradiciones culturales. Esta consideración ayudará a desarrollar un modelo más inclusivo y capaz de generar música que resuene con una audiencia global diversa.

Al abordar estos aspectos con detenimiento, no solo mejoraremos la capacidad y la eficacia de nuestro modelo de generación musical, sino que también navegaremos por el complejo panorama legal y ético asociado con el uso de datos musicales. Este enfoque reflexivo y meticuloso es fundamental para el éxito y la innovación sostenible en el campo de la inteligencia artificial aplicada a la música.

## **2.5. Trabajos relacionados**

En la sección de trabajos relacionados, se efectuará un análisis detallado de investigaciones anteriores y desarrollos tecnológicos en el ámbito de la generación de música con inteligencia artificial. Esta revisión crítica permitirá establecer el estado del arte y destacar las contribuciones del presente estudio. Los elementos a cubrir incluyen:

### **2.5.1. Revisión de la Literatura:**

- Modelos de Redes Neuronales Profundas: Investigaciones que utilizan RNNs (Redes Neuronales Recurrentes), CNNs (Redes Neuronales Convolucionales), y GANs (Redes Generativas Adversarias) para la generación de música.
  - DeepBach: a Steerable Model for Bach Chorales Generation, que utiliza modelos de aprendizaje profundo para generar corales en

el estilo de Bach.

- **Sistemas de Composición Asistida:** Herramientas que proporcionan asistencia a compositores y productores musicales, como Google's Magenta o IBM's Watson Beat.
  - **Magenta: Music and Art Generation with Machine Intelligence** que presenta una plataforma para la creación de música y arte mediante machine learning.
- **Aplicaciones de Modelado de Música:** Uso de modelos para tareas específicas como armonización automática, creación de acompañamiento y generación de melodías.
  - **Flow Machines: A New Generation of Music Composition Tools**, que explora el uso de IA para la composición de música popular.

### **2.5.2. Comparación de Modelos:**

Se evaluarán contra los modelos y sistemas existentes, identificando sus metodologías, capacidades, y limitaciones. Se prestará especial atención a:

- **Desempeño en Generación de Música:** Cómo los modelos existentes han logrado generar música que es tanto técnicamente sólida como estéticamente agradable.
  - **Comparación con AIVA**, un sistema de inteligencia artificial reconocido por la SACEM (Sociedad de Autores, Compositores y Editores de Música) como un compositor legítimo de música.

- Diversidad Estilística y Adaptabilidad: La habilidad de los modelos para generar música en varios estilos y su adaptabilidad a nuevos géneros.
  - Análisis de "Jukedeck", una start-up que desarrolló un sistema capaz de crear música en varios estilos y que fue adquirida por TikTok.
- Evaluaciones de Usuarios externos y Expertos: Las respuestas de músicos, compositores y el público general a las composiciones generadas por IA.
  - Estudios que utilizan encuestas y paneles de evaluación, como Evaluating Music Generation Systems with Steerable Metrics que proponen métricas específicas para la evaluación.

# Capítulo 3

## Diseño

En este proyecto, nos enfocamos en el uso de Redes Neuronales Recurrentes (RNN) para la generación de música a partir de entradas MIDI. Esta elección se debe a la capacidad inherente de las RNN para manejar secuencias de datos, lo cual es fundamental en el procesamiento de música, donde la temporalidad y el orden de las notas son esenciales.

### 3.1. Arquitectura

#### 3.1.1. Detalles de la Red Neuronal

La arquitectura que proponemos para la generación de música a partir de entradas MIDI se basa en el uso de Redes Neuronales Recurrentes (RNN), específicamente utilizando una configuración de capas LSTM (Long Short-Term Memory). Las RNN son ideales para procesar secuencias de datos, como las notas en una composición musical, debido a su capacidad para mantener un estado o memoria de los inputs anteriores. Sin embargo, las RNN tradicionales a menudo enfrentan desafíos, como el problema del desvanecimiento del gradiente, donde la influencia de la información de



entrada se pierde a medida que pasa a través de múltiples capas en una red profunda.

Las capas LSTM abordan este problema al introducir puertas que regulan el flujo de información. Estas puertas -puerta de olvido, puerta de entrada y puerta de salida- permiten que las LSTM conserven o descarten información a lo largo del tiempo, haciendo posible capturar dependencias a largo plazo en los datos. Esto es crítico en la música, donde la relación entre las notas no solo es inmediata sino que a menudo se extiende a lo largo de varias medidas.

Nuestra red empleará tres capas LSTM, cada una con 128 unidades. Esta configuración busca un equilibrio entre la capacidad de modelado y la eficiencia computacional. Con demasiadas unidades, la red podría volverse propensa al sobreajuste, aprendiendo peculiaridades del conjunto de datos de entrenamiento en lugar de generalizar patrones musicales. Con muy pocas, la red podría no tener suficiente capacidad para aprender la complejidad inherente a la música.

Para las capas intermedias, hemos seleccionado la función de activación ReLU (Rectified Linear Unit). Esta función tiene varias ventajas: es computacionalmente eficiente, lo que es crucial en el entrenamiento de redes profundas, y ayuda a mitigar el problema de las neuronas inactivas (también conocido como "muerte de ReLU") al proporcionar una activación no saturante. Esto significa que siempre que la entrada a la neurona es positiva, hay un gradiente para seguir durante el entrenamiento, lo cual es crucial para la actualización efectiva de los pesos en una red.

### 3.1.2. Manejo de Entradas y Salidas

El procesamiento de entradas MIDI para su uso en una red neuronal requiere una cuidadosa preparación de los datos para asegurar que la red pueda interpretar y generar efectivamente música. Las entradas MIDI, que son esencialmente instrucciones digitales para la reproducción de música,

incluyen información como el tono de las notas, su duración, la velocidad (intensidad) y posiblemente otros atributos como la instrumentación.

#### Preprocesamiento de Entradas MIDI

**Codificación One-Hot de Notas:** Cada nota en una pieza MIDI será convertida en un vector one-hot. En esta codificación, cada vector tiene una longitud igual al número de notas posibles (por ejemplo, 88 para un piano estándar), con un '1' en la posición que corresponde a la nota específica y '0' en todas las demás. Esta representación simplifica la entrada para la red, convirtiendo información compleja en un formato binario fácil de manejar.

**Manejo de la Duración y la Velocidad:** Además de la nota en sí, es importante considerar cómo representar la duración y la velocidad de cada nota. Esto podría implicar la inclusión de información adicional en el vector de entrada o el uso de una secuencia de vectores para representar una sola nota sostenida a lo largo del tiempo.

**Normalización:** Para mejorar la eficiencia del entrenamiento de la red, estos vectores podrían normalizarse para asegurar que los valores de entrada estén en un rango manejable para la red neuronal.

#### Generación de Salida y Conversión a MIDI

**Secuencia de Vectores de Salida:** La red generará una secuencia de vectores en respuesta a la entrada MIDI. Cada vector de salida corresponderá a una nota en la pieza musical generada, utilizando una codificación similar a la de las entradas.

**Decodificación de Vectores a Notas MIDI:** Los vectores de salida deben convertirse de nuevo en formato MIDI para su reproducción. Esto implica identificar la nota (o notas) que cada vector representa y reconstruir la información sobre la duración y la velocidad.

**Reconstrucción de la Pieza Musical:** Finalmente, la secuencia de notas MIDI decodificadas se ensamblará en una pieza musical coherente. Esto implica no solo colocar las notas en el orden correcto sino también ajustar la

duración y la velocidad para reflejar la interpretación deseada.

Este proceso de preprocesamiento y postprocesamiento es crucial para la efectividad de la red neuronal en la generación de música. La precisión en la conversión de datos MIDI a un formato que la red pueda entender y luego de vuelta a música reproducible es clave para el éxito del proyecto.

### 3.1.3. Regularización y Prevención del Sobreajuste

Para abordar el sobreajuste, una preocupación común en el entrenamiento de modelos de aprendizaje profundo, especialmente en redes complejas como las LSTM, implementaremos la técnica de dropout. El sobreajuste ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento, perdiendo así su capacidad para generalizar a nuevos datos. Esto es particularmente crítico en aplicaciones de generación musical, donde queremos que el modelo produzca música que sea variada y no simplemente una repetición de los ejemplos de entrenamiento.

#### Implementación de Dropout en las Capas LSTM

Tasa de Dropout del 20: Hemos decidido utilizar una tasa de dropout del 20, lo que significa que en cada paso del entrenamiento, un 20 de las unidades neuronales (o conexiones) en las capas LSTM se .apagarán. aleatoriamente. Este porcentaje es un equilibrio entre permitir que el modelo aprenda suficientes detalles de los datos y prevenir la dependencia excesiva de cualquier conjunto de características.

Efecto de Regularización: Dropout actúa como una forma de regularización. Al desactivar aleatoriamente diferentes neuronas durante el entrenamiento, se fuerza a la red a aprender patrones robustos que son generales a lo largo de diferentes subconjuntos de los datos. Esto ayuda a evitar que el modelo dependa demasiado de cualquier neurona o ruta específica, lo que puede llevar a un sobreajuste a los datos de entrenamiento.

Mejora de la Generalización del Modelo: Al reducir la complejidad del

modelo durante el entrenamiento, el dropout mejora la capacidad del modelo para generalizar a nuevos datos. Esto es esencial para la generación musical, donde deseamos que el modelo produzca piezas nuevas y creativas, no solo réplicas de lo que ha .escuchado”durante el entrenamiento.

**Equilibrio entre Aprendizaje y Generalización:** Es importante ajustar la tasa de dropout para encontrar el equilibrio correcto entre permitir que la red aprenda efectivamente de los datos de entrenamiento y mantener suficiente regularización. Una tasa demasiado alta puede impedir que la red aprenda adecuadamente, mientras que una tasa demasiado baja puede ser insuficiente para prevenir el sobreajuste.

### **3.1.4. Flujo de Datos y Procesamiento**

El flujo de datos comienza con la entrada MIDI, que se procesa en un formato adecuado para la red. Luego, esta secuencia procesada se alimenta a través de las capas LSTM. La red genera una secuencia de vectores de salida que se convierten de nuevo en un formato MIDI para crear la pieza musical.

## **3.2. Soluciones Alternativas**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.