



# Generación de música con modelo Transformer

Felipe Andres Castillo

email: 3.1416.p@ciencias.unam.mx

*Diplomado Inteligencia Artificial y Ciencia de Datos*

25 DE ABRIL DE 2025



## Resumen

La música es uno de los grandes placeres de la vida y está profundamente ligada a nuestra identidad y experiencia. Me encanta la música y siento una profunda admiración por quienes se dedican a componer, crear e interpretar. Aunque disfruto mucho de ella, siempre la he vivido desde el rol de espectador, como parte del público. Por eso, la posibilidad de utilizar la inteligencia artificial para intentar imitar el acto creativo de componer música, sin tener el talento y el conocimiento técnico, me resulta especialmente interesante.

En este proyecto se implementó un modelo de generación de melodías, usando una arquitectura *Decoder-only Transformer*. El modelo es entrenado con los datasets MAESTRO (<https://magenta.tensorflow.org/datasets/maestro>) y GiantMIDI (<https://github.com/bytedance/GiantMIDI-Piano>). En el proceso se explora lo que es la *tokenización* de archivos MIDI, el mecanismo de *atención global relativa* (relative global attention) y el almacenamiento de memoria caché para mejorar los tiempos de inferencia.

Para llevar a cabo este último objetivo, fue imprescindible comprender en profundidad el funcionamiento del modelo, tanto desde una perspectiva teórica como a nivel de implementación en código. Si bien la integración se tradujo en una mejora en el tiempo de inferencia, esta fue de apenas unos pocos segundos. En cuanto a los resultados, las melodías generadas me llevan a una conclusión clara: imitar la creatividad y la sensibilidad que implica componer música es, y probablemente seguirá siendo, una tarea sumamente compleja.

En el notebook he detallado completamente sobre el modelo, añadiendo texto y comentarios al código. Si se quiere ejecutar el modelo, he adjuntado el archivo del modelo entrenado, y dos archivos correspondientes al preprocesamiento de los datos (notes\_train.npz y notes\_test.npz), pues ejecutar desde cero toma bastante tiempo. Por cierto, he utilizado Google Colab y la GPU A100.

Este trabajo está basado en el notebook de Keras ([https://keras.io/examples/generative/midi\\_generation\\_with\\_transformer/](https://keras.io/examples/generative/midi_generation_with_transformer/)), que a su vez está basado en el artículo "Music Transformer", por Huang et al. (2018) (<https://arxiv.org/abs/1809.04281>).