

INTRODUCCIÓN A LOS SISTEMAS DE PERCEPCIÓN

BLOQUE: PERCEPCIÓN VISUAL Y AUDITIVA

By PhD. Raúl Llasag Rosero

raulhomero.llasag-externo@unir.net

GitHub: [rhllasag](https://github.com/rhllasag)

9/11/2025



Presiona **Espacio** o → para navegar



AGENDA

- ▶ Revisión de la Visión Artificial
- ▶ Principales descubrimientos en audio
- ▶ Temas de investigación actuales



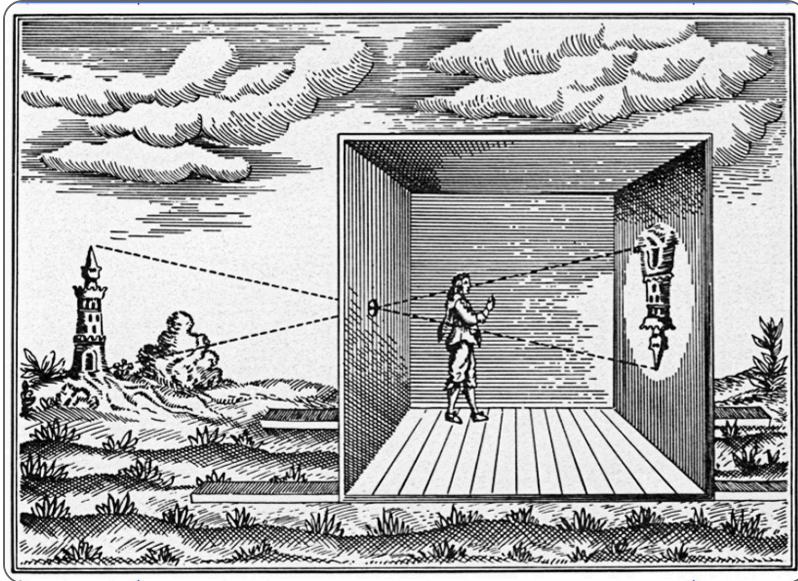
PRINCIPALES EVENTOS EN VISIÓN ARTIFICIAL



SIGLO IV A.C.

Cámara Oscura - Aristóteles

"La luz que pasa a través de un pequeño agujero crea una imagen invertida en la superficie opuesta."



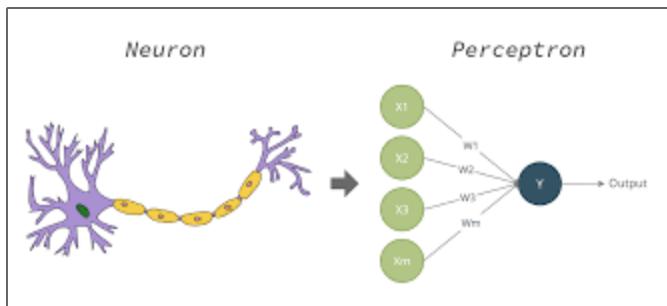
<https://www.photoion.co.uk/blog/camera-obscura>

Presiona ↓ para continuar

 1957

Perceptrón - Frank Rosenblatt

"Tipo de red neuronal artificial que actúa como un clasificador binario básico, inspirado en la función biológica de las neuronas."



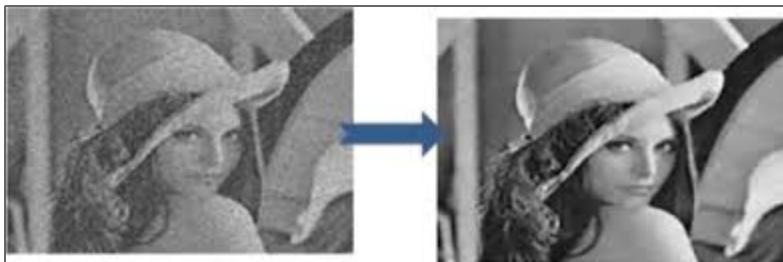
<https://news.cornell.edu>

Presiona ↓ para continuar



Procesamiento Digital de Imágenes - Azriel Rosenfeld

"El procesamiento digital de imágenes es el uso de una computadora digital para procesar imágenes digitales a través de un algoritmo. Se utiliza para mejora de imagen, restauración, compresión y análisis."



Eliminación de ruido en imágenes corrompidas con ruido impulsivo, gaussiano o una mezcla de ruido impulsivo y gaussiano - Ali Awad

Presiona ↓ para continuar



AÑOS 60

Visión por Computadora - Larry Roberts

"La visión por computadora es el campo que se ocupa de cómo las computadoras pueden obtener comprensión de alto nivel a partir de imágenes digitales. Desde reconstrucción de escenas 2D a 3D, reconocimiento de objetos y restauración de imágenes."

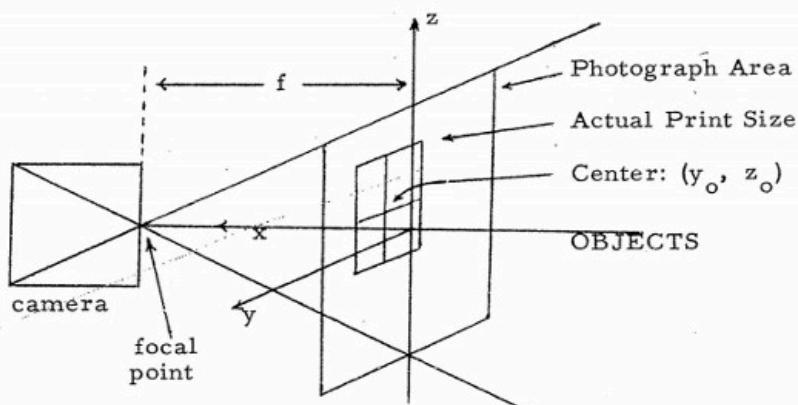


Figure 1:

Camera Transformation

<https://zbigatron.com/the-early-history-of-computer-vision/>

Presiona ↓ para continuar



Detección de Bordes - John Canny

"La detección de bordes es una técnica fundamental de procesamiento de imágenes que identifica puntos en una imagen digital donde el brillo cambia abruptamente. Estos puntos típicamente representan límites de objetos, marcas superficiales o cambios de iluminación, haciendo que la detección de bordes sea crucial para la extracción de características y el reconocimiento de objetos en visión por computadora."



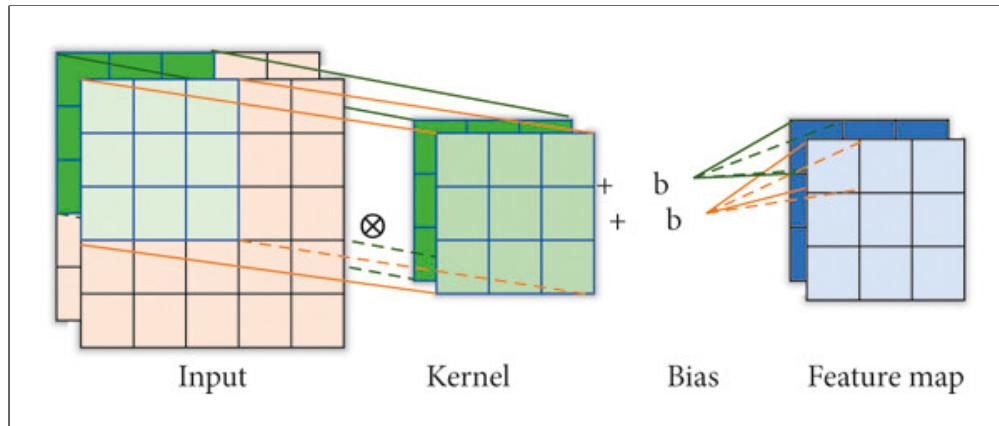
https://en.wikipedia.org/wiki/Edge_detection

Presiona ↓ para continuar



Redes Neuronales Convolucionales - Yann LeCun

"Las operaciones de convolución son particularmente adecuadas para procesar datos que tienen una topología similar a una cuadrícula."



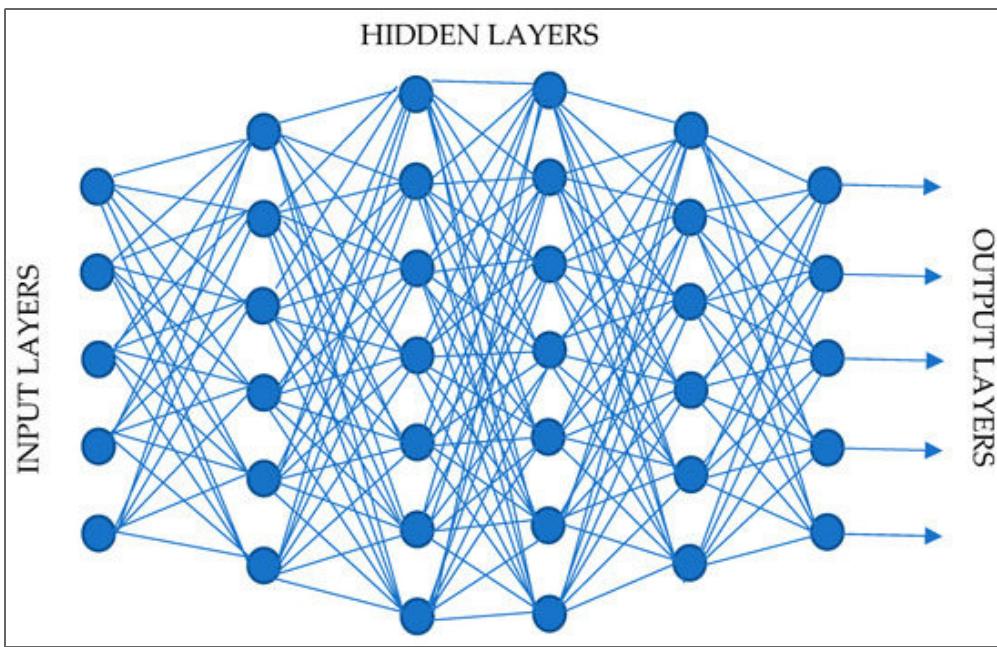
<https://onlinelibrary.wiley.com/doi/10.1155/2021/5573031>

Presiona ↓ para continuar



Revolución del Aprendizaje Profundo - Geoffrey Hinton, Yoshua Bengio, Yann

"Las Redes Neuronales Profundas (DNN) son una clase de algoritmos de aprendizaje automático que utilizan múltiples capas de transformaciones no lineales para aprender patrones complejos en los datos."



[Una Revisión de los Algoritmos de Aprendizaje Profundo y sus Aplicaciones en Salud - Hussein Abdel-Jaber y et al.](#)

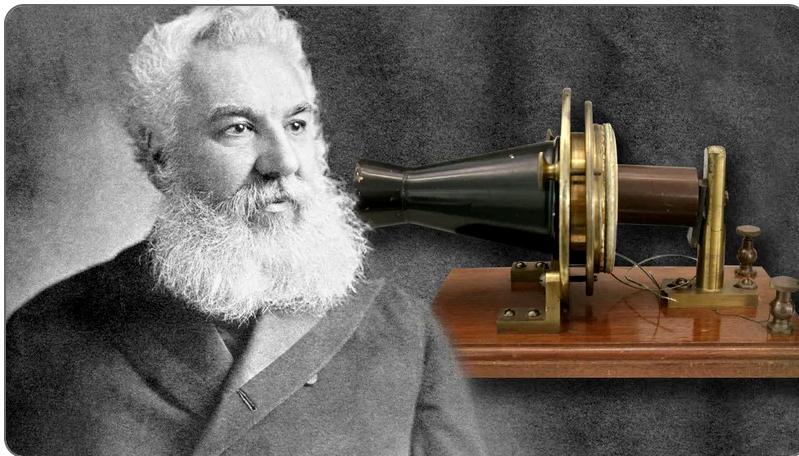
Presiona → para finalizar

🎵 PRINCIPALES DESCUBRIMIENTOS EN AUDIO

♫ 1876

Teléfono - Alexander Graham Bell

"La primera transmisión exitosa del habla humana a través de cables eléctricos, sentando las bases para la tecnología de comunicación de audio."



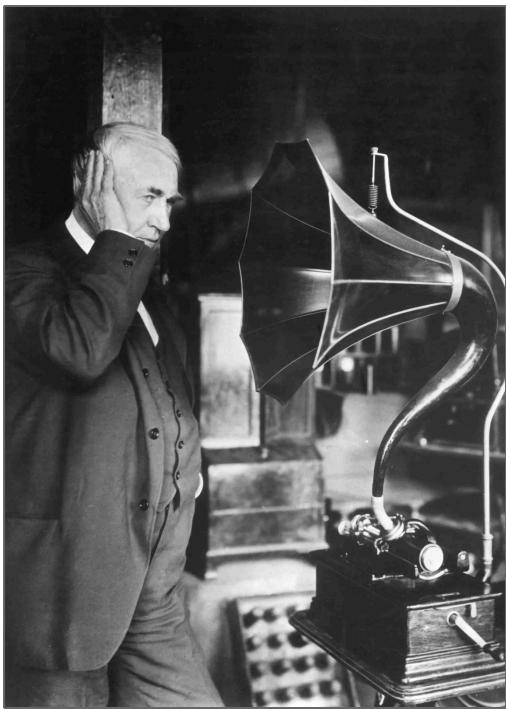
<https://www.history.com/topics/inventions/telephone>

Presiona ↓ para continuar



Fonógrafo - Thomas Edison

"El primer dispositivo capaz de grabar y reproducir sonido, revolucionando la forma en que los humanos podían capturar y reproducir audio."



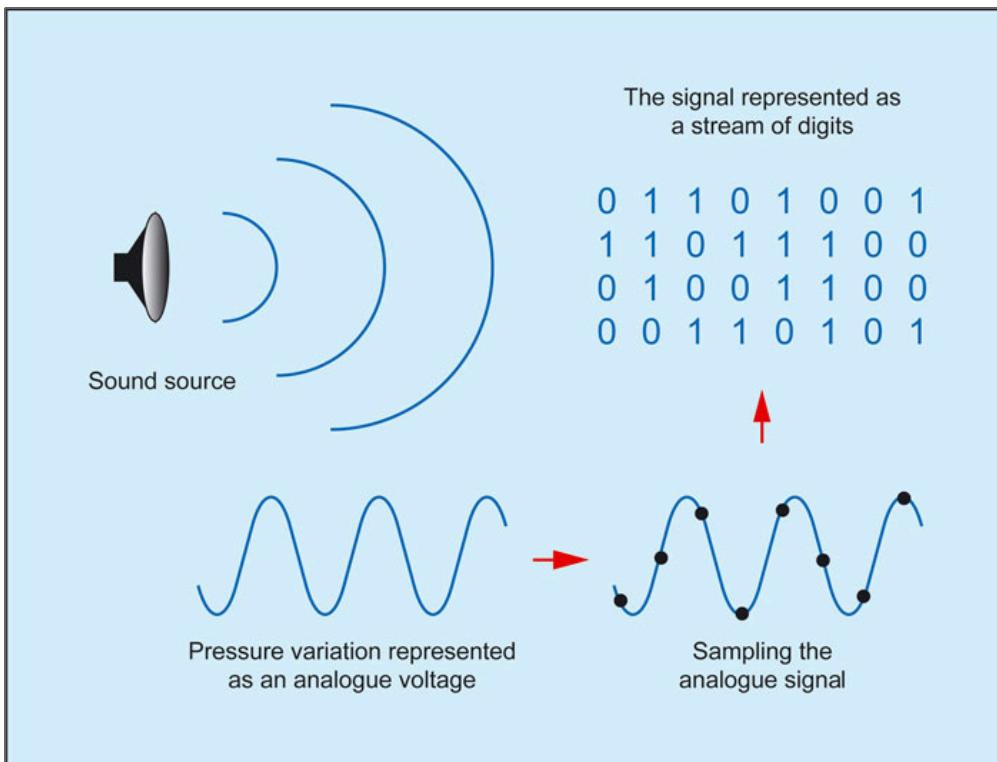
<https://www.britannica.com/biography/Thomas-Edison>

Presiona ↓ para continuar



Audio Digital - Alec Reeves (PCM)

"La Modulación por Código de Pulsos (PCM) introdujo el concepto de convertir señales de audio analógicas a formato digital, permitiendo reproducción y procesamiento de audio preciso."



<https://www.soundonsound.com/techniques/digital-myth>

Presiona ↓ para continuar

 **AÑOS 50**

Reconocimiento de Voz - Bell Labs

"Sistemas tempranos de reconocimiento de voz capaces de reconocer dígitos hablados, marcando el comienzo de la tecnología de reconocimiento automático de voz."



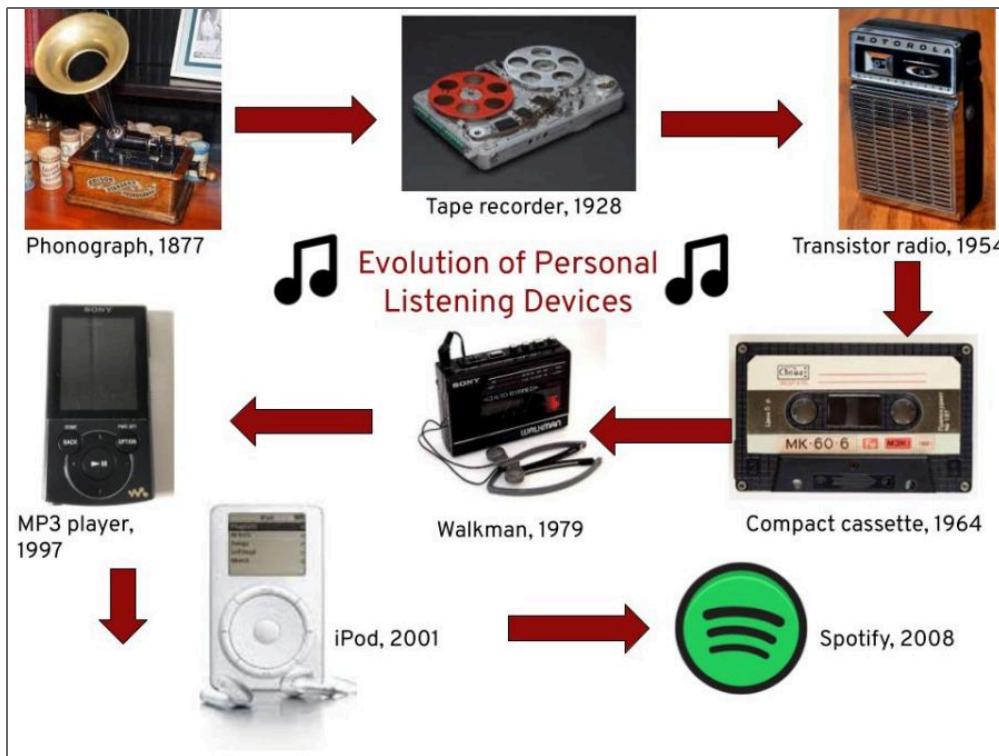
<https://embeddedcomputing.com/technology/ai-machine-learning>

Presiona ↓ para continuar



Compresión MP3 - Karlheinz Brandenburg

"MPEG-1 Audio Layer III revolucionó el audio digital al permitir compresión de sonido de alta calidad, haciendo que la distribución de música digital fuera práctica y generalizada."



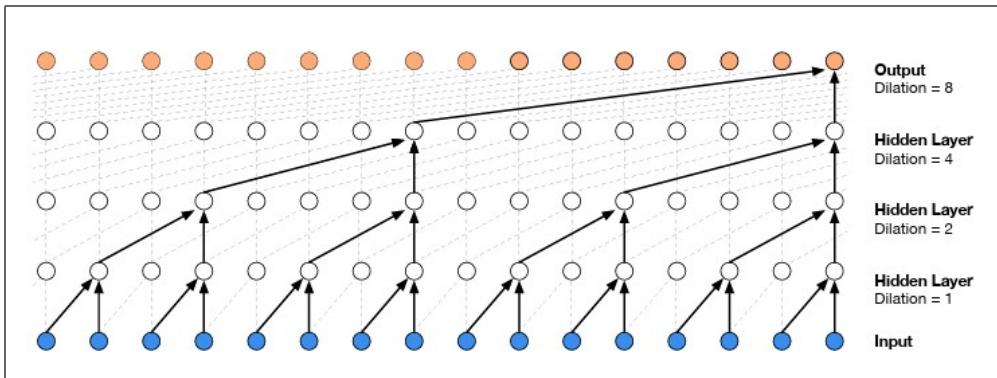
[Del pop al punk: perfeccionando el juego de las listas de reproducción - Emma Riddell](#)

Presiona ↓ para continuar



Aprendizaje Profundo para Audio - Hinton, Deng, y otros

"Las redes neuronales profundas transformaron el procesamiento de audio, permitiendo avances revolucionarios en reconocimiento de voz, generación de música y síntesis de audio."



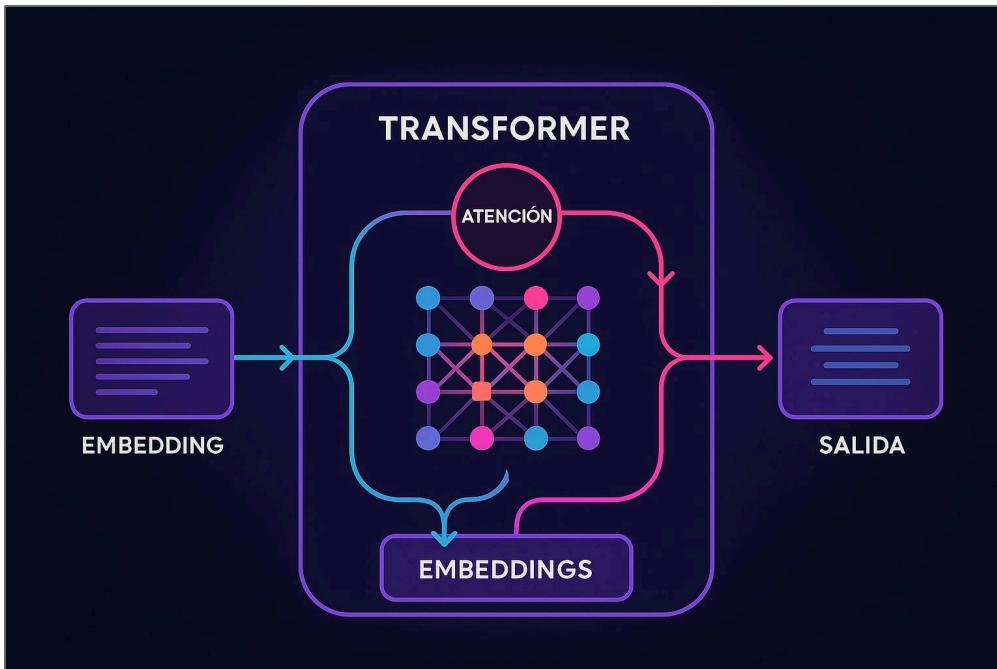
WaveNet: Un Modelo Generativo para Audio Crudo - Aaron van den Oord y et al.

Presiona ↓ para continuar



Modelos Transformer para Audio - OpenAI, Google, Meta

"Las arquitecturas Transformer como Whisper, WaveNet y AudioLM han logrado rendimiento a nivel humano en reconocimiento de voz, síntesis y generación de música."



Transformers explicados fácil: la base de la comprensión en IA

Presiona → para finalizar



TEMAS DE INVESTIGACIÓN ACTUALES



VISIÓN POR COMPUTADORA

- ▶ Transformers de Visión (ViTs)
- ▶ Detección de Objetos 3D
- ▶ Campos de Radiancia Neural (NeRF)
- ▶ IA Multimodal



PROCESAMIENTO DE AUDIO

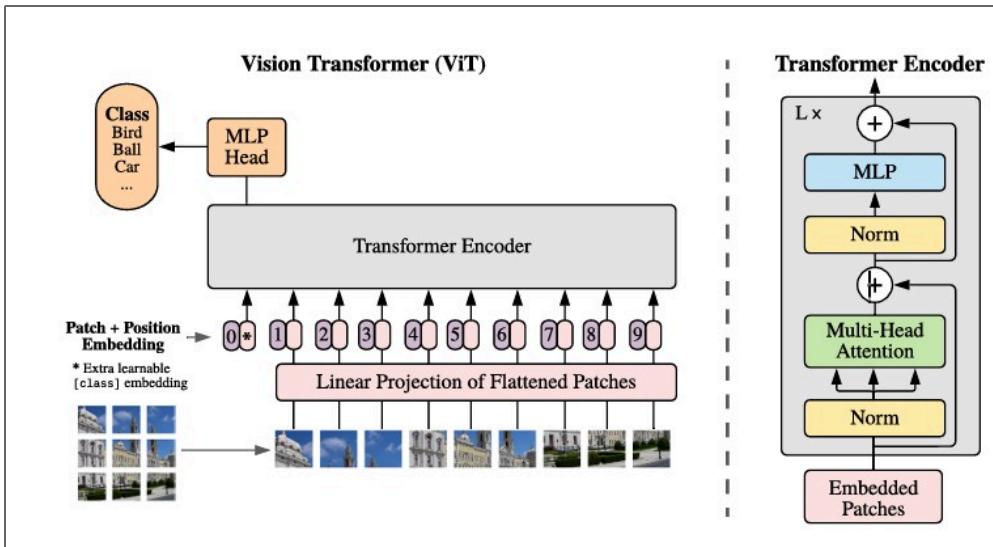
- ▶ Modelos Fundacionales de Audio
- ▶ Traducción de Voz en Tiempo Real
- ▶ Síntesis de Audio Neural
- ▶ Aprendizaje Audio-Visual

Presiona ↓ para explorar cada tema



TRANSFORMERS DE VISIÓN (ViTs)

"Los Vision Transformers aplican la arquitectura transformer directamente a parches de imagen. A diferencia de las Redes Neuronales Convolucionales (CNN) tradicionales que usan campos receptivos locales, los ViTs utilizan auto-atención para modelar dependencias globales de largo alcance entre diferentes parches de imagen, proporcionando una mejor comprensión del contexto general de la imagen."



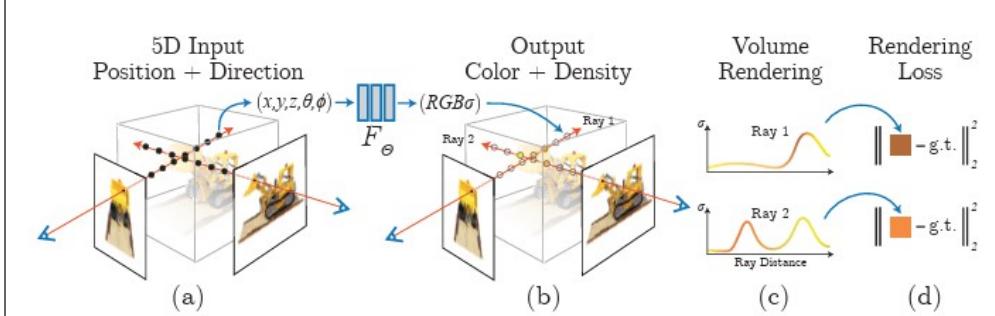
Referencias:

- Dosovitskiy, A., et al. (2020). "Una Imagen Vale 16x16 Palabras: Transformers para Reconocimiento de Imágenes a Escala." *arXiv:2010.11929*
- Liu, Z., et al. (2021). "Swin Transformer: Transformer de Visión Jerárquico usando Ventanas Desplazadas." *ICCV 2021*
- Presiona ↓ para continuar



CAMPOS DE RADIANCIA NEURAL (NERF)

"NeRF representa escenas como funciones 5D continuas que mapean coordenadas espaciales y direcciones de visualización a color y densidad, permitiendo síntesis fotorrealista de vistas novedosas."



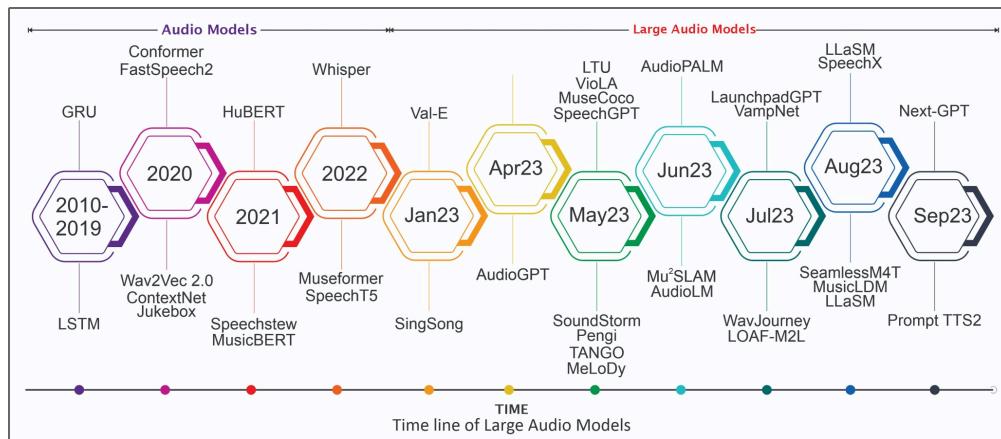
Referencias:

- Mildenhall, B., et al. (2020). "NeRF: Representando Escenas como Campos de Radiancia Neural para Síntesis de Vistas." *ECCV 2020*
- Barron, J. T., et al. (2021). "Mip-NeRF: Una Representación Multiescala para Campos de Radiancia Neural Anti-Aliasing." *ICCV 2021*

Presiona ↓ para continuar

MICROPHONE MODELOS FUNDACIONALES DE AUDIO

"Modelos pre-entrenados a gran escala como Whisper, AudioLM y MusicLM están revolucionando la comprensión de audio, generación y aplicaciones cross-modales."



<https://github.com/EmulationAI/awesome-large-audio-models>

Referencias:

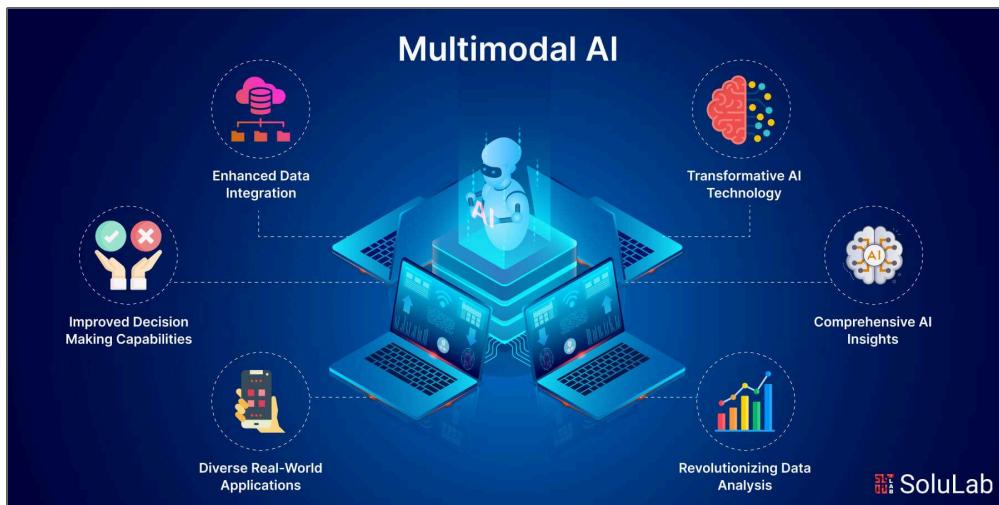
Borsos, Z., et al. (2023). "AudioLM: un Enfoque de Modelado de Lenguaje para Generación de Audio." *IEEE/ACM Transactions on Audio, Speech, and Language Processing*

Presiona ↓ para continuar



IA MULTIMODAL

"La integración de modalidades de visión y audio permite sistemas de IA más robustos y conscientes del contexto, desde comprensión de video hasta IA encarnada e interacción humano-computadora."



Referencias:

Alayrac, J. B., et al. (2022). "Flamingo: un Modelo de Lenguaje Visual para Aprendizaje Few-Shot." *NeurIPS 2022*

Presiona → para finalizar

¿CÓMO PODEMOS OPTIMIZAR LA INTEGRACIÓN DE LOS MECANISMOS DE CAPTACIÓN DE SEÑALES AUDITIVAS Y VISUALES HUMANAS EN EL DISEÑO DE SISTEMAS DE PERCEPCIÓN COMPUTACIONAL?

By PhD. Raúl Llasag Rosero

raulhomero.llasag-externo@unir.net

GitHub: [rhllasag](#)

LinkedIn: [raúl-homero-llasag-rosero-992574135](#)

