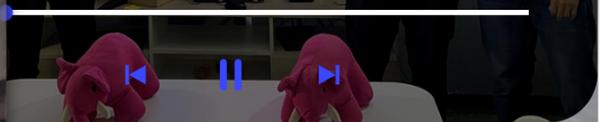
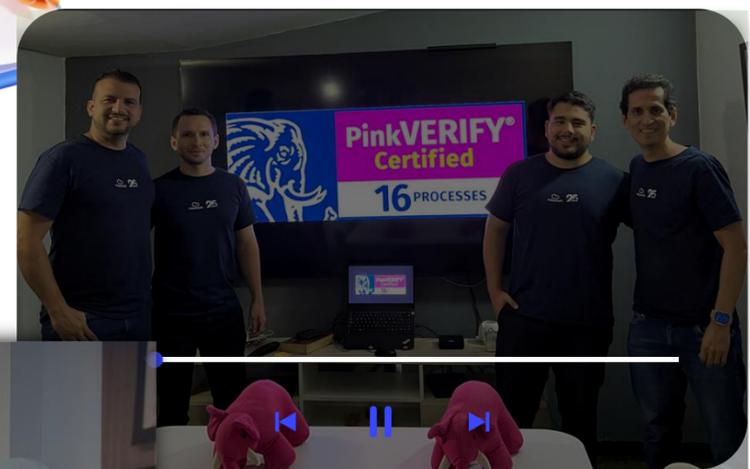
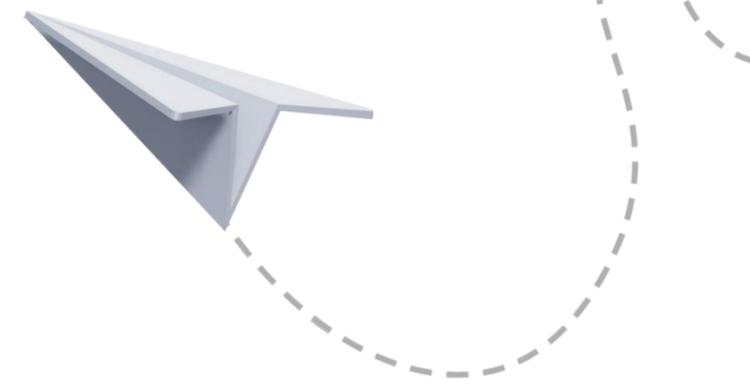


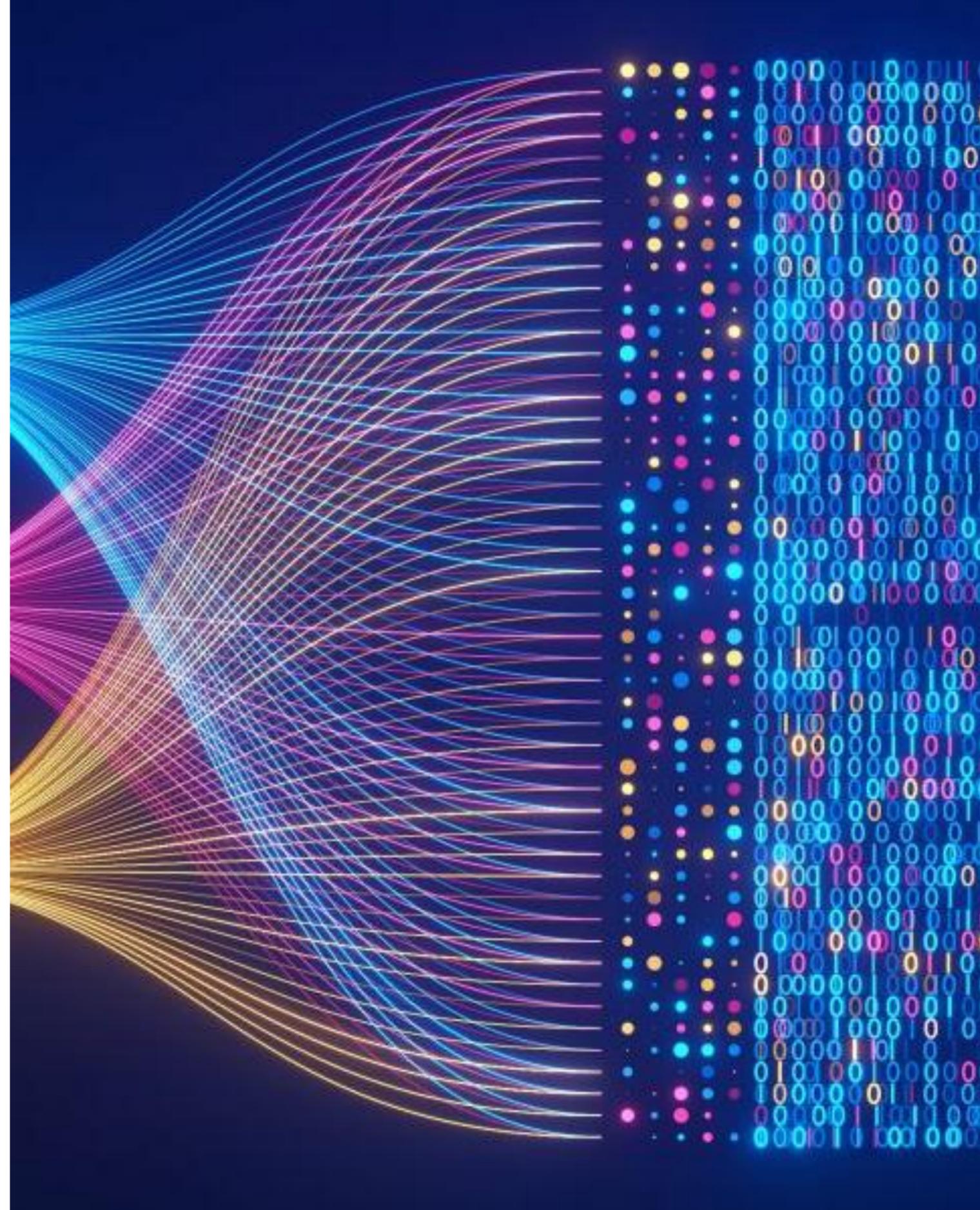
Deep Learning en voz y texto

Bloque 4



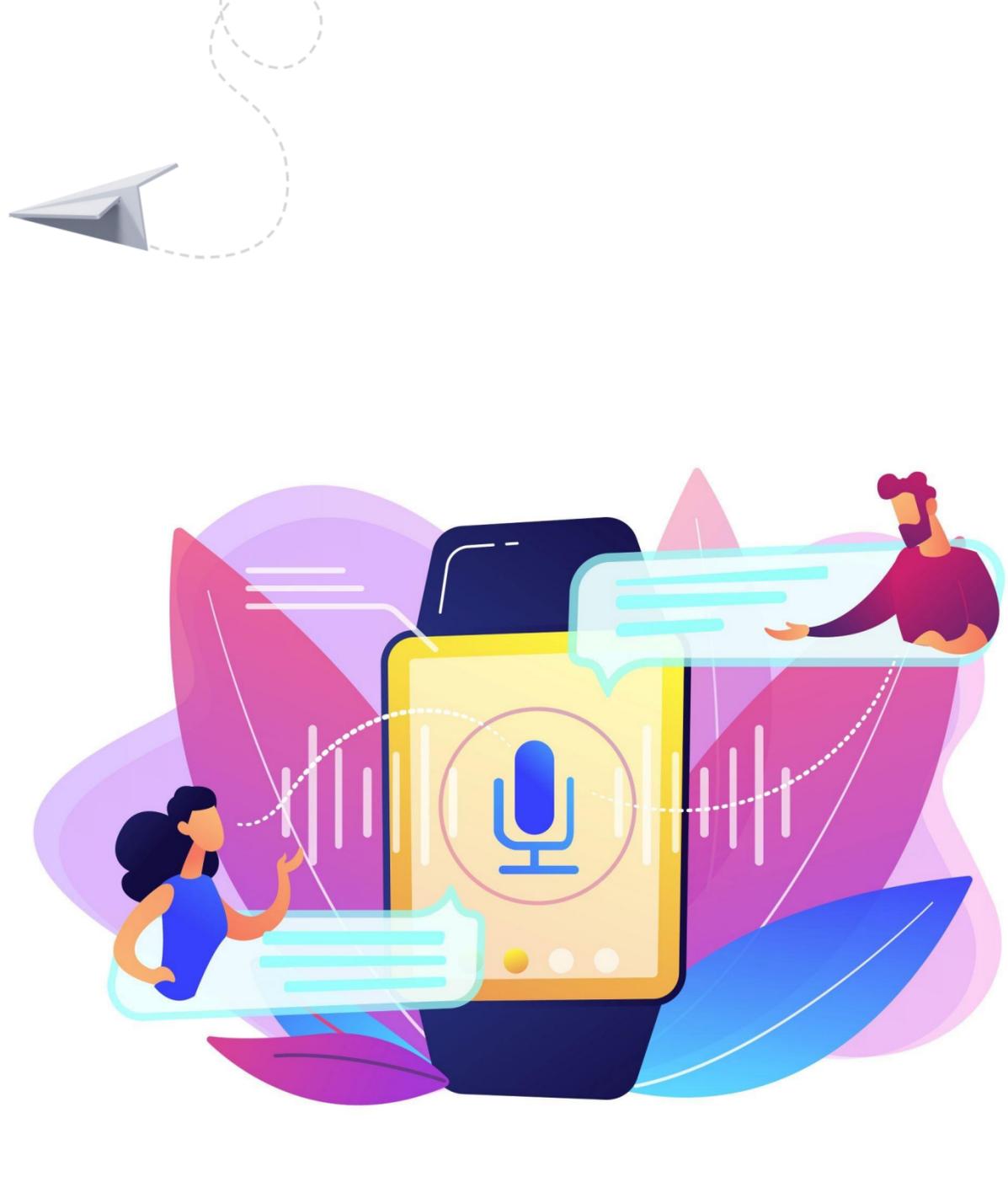
Deep Learning:

Del sonido al significado: cómo las máquinas entienden y generan lenguaje



¿Qué es el procesamiento de voz y por qué es complejo?

Entender la voz humana implica decodificar una señal variable afectada por acentos, ruido y emociones. Las máquinas deben aprender a aislar palabras útiles de este flujo sonoro complejo, transformando ondas acústicas en significado, algo que para nosotros es natural pero que para la IA representa un desafío técnico enorme.





De sonido a espectrograma: cómo “escucha” una red

Las redes no procesan sonido directamente: lo transforman en espectrogramas (mapas visuales de frecuencia vs tiempo). Al tratarlos como imágenes, pueden identificar patrones equivalentes a letras, palabras y hasta emociones en el habla.



01. Fonemas

Unidades básicas de sonido

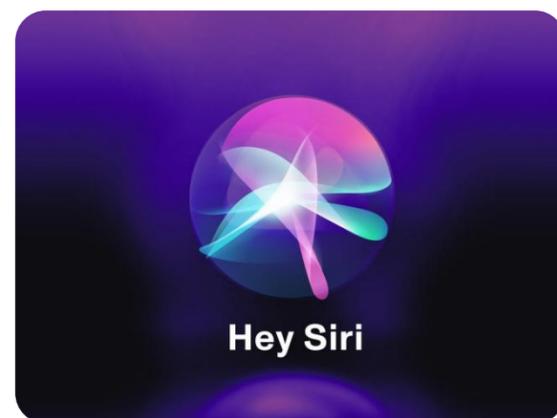
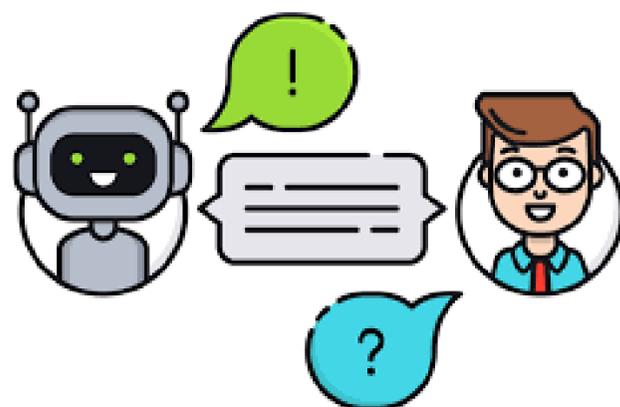
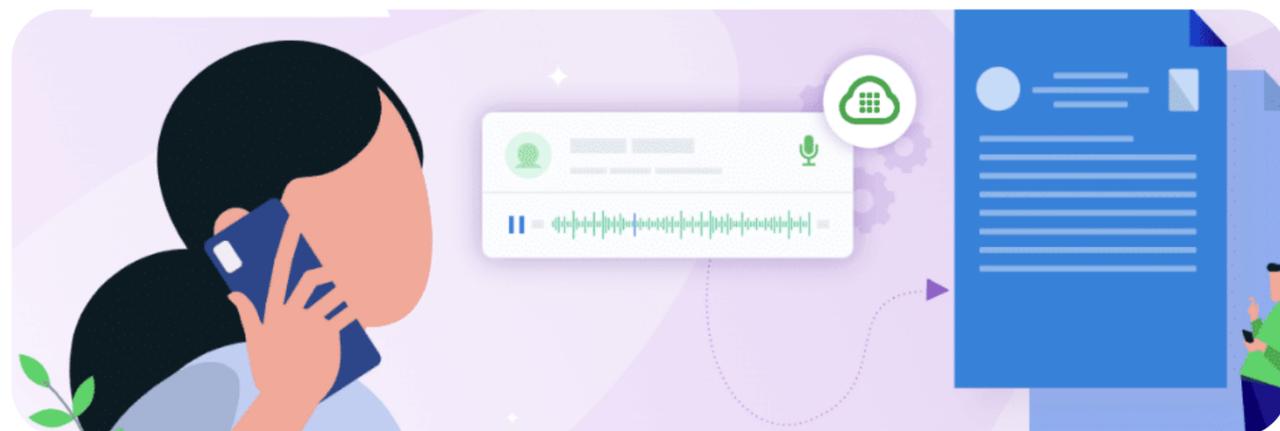


02. Transiciones entre palabras



03. Características emocionales





Aplicaciones Prácticas

- **Asistentes virtuales:** Siri, Alexa, Google Assistant
- **Reconocimiento de emociones** en llamadas o videojuegos
- **Transcripción automática** de audio a texto
- **Traducción simultánea**
- **Control por voz** en dispositivos inteligentes

¿Qué es el lenguaje natural (NLP)?

NLP = Natural Language Processing

El Procesamiento del Lenguaje Natural (NLP) aplica redes neuronales para modelar el lenguaje humano, superando sus retos únicos (polisemia, ironía, contextos culturales) mediante la fusión de lingüística computacional, estadística bayesiana y arquitecturas como Transformers, logrando tareas desde traducción automática hasta análisis de sentimientos.



Procesamiento de Texto para Redes Neuronales

El texto se transforma en representaciones numéricas llamadas embeddings (vectores densos que codifican significado). Estas representaciones:



01. Traducen palabras a números

Cada término se mapea a un vector en espacio multidimensional



02. Capturan relaciones semánticas

Operaciones vectoriales reflejan lógica lingüística
 $\text{rey} - \text{hombre} + \text{mujer} \approx \text{reina}$



03. Preservan contexto

Palabras similares o usos equivalentes tienen vectores cercanos en el espacio.

Riesgos y retos del DL en voz y texto

El DL en voz y texto no solo revoluciona la comunicación, sino que también introduce riesgos como deepfakes, desinformación y sesgos algorítmicos. Estos desafíos técnicos y éticos exigen marcos de control responsables para evitar que el progreso tecnológico comprometa la verdad y la equidad.



Desinformación generada por IA

Sesgos lingüísticos
Resultados injustos



Deepfakes de voz

Imitación realista de voces humanas.

Muchas Gracias

Valeria Rudas Ruiz