

Totoso, Sototo, Traduki kai Babilo.



Parecen el nombre de unos payasos esperantistas al estilo de “Gavi, Fofo , Miliki y Fofito’. Pero son los nombres ficticios que he puesto a las tecnologías de voludi.

- Totoso: (TTS: Text To Speech): Convierte texto en sonido.
- Sototo: (STT : Speech To Text): Convierte sonido de voz a Texto.
- Traduki: Traduce un texto de un lenguaje a otro.
- Babilo: Con el que se puede chatear en esperanto.

Al igual que ocurría con los payasos, uno parece listo y los otros tontos. Pero todos hacen su papel de forma inteligente, usando modelos de IA.

Un modelo IA de este tipo almacena palabras en forma de vectores y matrices de n dimensiones, llamadas tensores. Imagina un cubo de 3 dimensiones con muchas flechas o vectores en su interior que representan palabras. La posición, tamaño, dirección lo relaciona con las otras palabras. Cuantas más palabras, parámetros y relaciones, más grande son los archivos de datos.

Modelos más grandes pueden hacerlo mejor, pero usan más memoria y son más lentos. Por eso usan ordenadores muy grandes o con sofisticados chips.

Como curiosidad, uno de los chips más avanzados y rápidos con más de 1000 núcleos y especializado en IA es de la empresa Esperanto.ia



<https://www.esperanto.ai/>

La intención es usar proyectos y modelos abiertos ligeros que puedan instalarse en ordenadores personales o incluso en el móvil. Y Combinarlos para conseguir una comunicación multilingüe, pero especialmente enfocada al Esperanto.

Componiendo a nuestros amigos payasos en una cadena de acciones se pueden crear muchos escenarios de diálogo.

Un escenario simple es que hables en Castellano y se repita lo que dices en otro idioma por ejemplo Esperanto. Sería un simple asistente de traducción.

SONIDO → Sototo → TEXTO → Traduki → TEXTO → Totoso → SONIDO



Sototo → Traduki → BABILO → Traduki → Totoso

Escenarios más complejos permiten hablar con alguien en esperanto, incluso si uno de los dos no sabe. Poniendo la capa de traducción en el sitio apropiado.

Un caso especial es que el interlocutor sea una IA. Aunque los modelos como GPT3.5 que son relativamente pequeños ya hablan bastante bien esperanto, hay otros que no.

Totoso(TTS: Text To Speech):

Es el sistema que convierte texto en sonido. Solía llamar mucho la atención en el pasado, ya que son tecnologías antiguas.

Se trata de que se pueda entender, no en buscar la entonación perfecta. Aunque ya hay modelos de IA que pueden imitar la entonación de cualquier persona, o incluso cantar.

Como queremos soluciones sencillas, usar el TTS del móvil es lo más práctico.

Es bastante sencillo implementar un TTS con los servicios y programas instalados en los propios móviles. Eliges el idioma , le indicas el texto y se genera el sonido.



El problema es que de momento no hay idioma Esperanto como parte del móvil.

Estaba la alternativa de usar un servicio remoto que hiciera el Speech. Hice pruebas con varios y no estaba mal el resultado. Pero usan el truco de hacer una transcripción fonética al Polaco y pasaban el resultado a un Speech Polaco estándar.

Como hay proyectos abiertos que hacen esa transcripción, con los datos y reglas que tome de ellos hice la implementación en Delphi para Voludi.

Ahora puede hacerse el TTS en el propio móvil, haciendo esa transcripción del texto esperanto y usar el speech estándar de Polaco que sí puede instalarse fácilmente. No es perfecto pero es entendible.

SoToTo 'escucha' el sonido y trata de identificar palabras.



Tiene los mismos problemas que tendrías tú para transcribir un audio a texto.

Conocer el idioma que se habla, sería la primera condición. Si no tienes nociones de una lengua, no la puedes transcribir a texto.

Incluso si eres bilingüe o multilingüe, ayudaría saber de antemano el idioma del audio para poder transcribir correctamente desde la primera palabra.

Hay que tener en cuenta los localismos de cada idioma. Por ejemplo en el Castellano el ceceo y seseo pueden transcribirse incorrectamente. También es importante un audio de calidad y una correcta vocalización de las palabras.

Los propios móviles y sistemas operativos ya incorporan sistemas STT muy avanzados en casi todos los idiomas. Pero de momento, el esperanto no está entre ellos.

Hay un proyecto abierto llamado VOSK que está en pleno desarrollo y expansión. Y que tiene modelos de Esperanto.

- Vosk ofrece dos tipos de modelos: **pequeños y grandes**.
- Los **modelos pequeños** son ideales para tareas limitadas en aplicaciones móviles y aplicaciones de escritorio. Tienen un tamaño de aproximadamente **50 MB** y requieren alrededor de **300 MB** de memoria en tiempo de ejecución.
- Los **modelos grandes** se utilizan para transcripciones de alta precisión en servidores. Estos modelos aplican algoritmos de IA avanzados y requieren hasta **16 GB** de memoria. [Se recomienda ejecutarlos en servidores potentes, como máquinas i7 o AMD Ryzen1.](#)


Hay 2 posibles soluciones para incorporar Vosk a una aplicación móvil.

- 1) De forma nativa usando un modelo pequeño y la biblioteca de sistema vosk.so. En total unos 60 MB
- 2) Usando un servidor remoto que transcriba

Traduki.

Traduce un texto de un lenguaje a otro. Es un elemento central



[LibreTranslate - API de traducción automática gratuita y de código abierto](#) 
[Komputeko](#)

Babilo:

Es un sistema con el que se puede chatear en esperanto con cualquier sistema o persona.



Tú hablas en esperanto, Sototo y traduki lo pasan al idioma de tu interlocutor, que le llega el texto y responde, traduki lo traduce de nuevo a esperanto y Totoso lo convierte en sonido.

Supongamos que tu no hablas esperanto y tu interlocutor si ., usando las mismas conversiones de forma inversa, consigues chatear con el.

Un uso interesante es hablar con sistemas como ChatGPT en esperanto. Hay muchos modelos LLM entrenados en muchos idiomas diferentes. Poder hablar con ellos en esperanto puede ser divertido.