

Parte 1

Azure cognitive Services

Este apartado tendrá más que el resto por el especial interés profesional en las tecnologías de Azure.

Microsoft Azure Cognitive Services es una colección de APIs alimentadas por IA desarrolladas por Microsoft que permiten a los desarrolladores añadir capacidades cognitivas a sus aplicaciones sin tener que tener experiencia directa en el campo inteligencia artificial o el datascience. Estas APIs cubren visión, voz, lenguaje, búsqueda y conocimiento, permitiendo la creación de aplicaciones más inteligentes e interactivas. A continuación, nos centraremos en las tecnologías que involucran al NLP

Servicios de Voz de Azure

Speech to Text / Voice to Text (STT/VT)

Descripción:

El servicio de reconocimiento de voz de Azure convierte audio en texto en tiempo real. Este servicio puede ser utilizado para transcribir llamadas, reuniones, y otros tipos de audio en texto legible y editable.

Características:

- **Precisión y Personalización:** El servicio se puede personalizar con modelos de lenguaje específicos del dominio y con voces personalizadas para mejorar la precisión.
- **Multi-idioma:** Soporta múltiples idiomas, incluyendo español, y puede reconocer varios dialectos y acentos.
- **Contexto:** Utiliza el contexto del discurso para mejorar la precisión de la transcripción.

Usos Comunes:

- Transcripción de reuniones.
- Subtítulos automáticos para videos.
- Asistencia en centros de llamadas.

Ej de uso: <https://www.youtube.com/watch?v=Zmyp1XHgyjw>

Text to Speech (TTS)

Descripción:

Convierte texto en habla realista utilizando modelos neuronales avanzados. Azure TTS ofrece voces que suenan naturales y pueden expresar emociones y entonaciones.

Características:

- **Neural TTS:** Utiliza redes neuronales profundas para generar una voz más natural y expresiva.
- **Personalización:** Los usuarios pueden crear voces personalizadas para sus aplicaciones.
- **Multi-idioma y Multi-dialecto:** Admite varios idiomas y dialectos, incluyendo voces específicas en español.

Usos Comunes:

- Lectura automatizada de texto para aplicaciones de accesibilidad.
- Generación de voces para asistentes virtuales.
- Creación de contenido de audio como audiolibros y podcasts.

Guías:

Español - https://www.youtube.com/watch?v=Xe9mo_lexZ0

Ingles - <https://www.youtube.com/watch?v=dl0amatX5zs>

Speech Translation

Descripción:

Ofrece traducción en tiempo real de voz a voz y de voz a texto, permitiendo una comunicación fluida entre hablantes de diferentes idiomas.

Características:

- **Traducción en Tiempo Real:** Permite la traducción instantánea de conversaciones en tiempo real.
- **Multi-idioma:** Soporta numerosos pares de idiomas, facilitando la comunicación global.
- **Contextualización:** Utiliza el contexto del discurso para mejorar la precisión de la traducción.

Usos Comunes:

- Comunicación en conferencias y reuniones internacionales.
- Asistencia en viajes y turismo.
- Integración en aplicaciones de mensajería y comunicación.

Arquitectura y Tecnología

1. Modelos Neuronales: Utiliza modelos de redes neuronales profundas que se entrenan con grandes conjuntos de datos para aprender las sutilezas del lenguaje humano.
2. Servicios en la Nube: Se ejecuta en la nube de Azure, lo que permite escalabilidad y accesibilidad global.
3. APIs y SDKs: Proporciona APIs y SDKs para varios lenguajes de programación, facilitando la integración en diversas aplicaciones.
4. Seguridad y Privacidad: Implementa altos estándares de seguridad y privacidad, protegiendo los datos de los usuarios y cumpliendo con las normativas internacionales.

Web de interés y guías:

Choose an Azure Cognitive Services technology

<https://learn.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/cognitive-services>

Landing Azure AI Services

<https://azure.microsoft.com/es-es/products/ai-services>

<https://azure.microsoft.com/en-us/products/ai-services>

Soluciones escalables con Cognitive services

<https://azure.microsoft.com/es-es/blog/azure-cognitive-services-for-building-enterprise-ready-scalable-ai-solutions/>

Pequeño blog de la tecnología (recomendable de leer)

<https://www2.deloitte.com/es/es/pages/technology/articles/introduccion-a-los-servicios-cognitivos-de-azure.html>

Amazon Alexa

Amazon Alexa es un asistente virtual que utiliza tecnologías avanzadas de reconocimiento de voz y aprendizaje profundo para interactuar con los usuarios. Alexa permite la creación de "skills" (aplicaciones) personalizadas mediante el uso de herramientas de desarrollo proporcionadas por Amazon

Nuance Communications

Nuance es pionera en el desarrollo de tecnologías de reconocimiento de voz y procesamiento de lenguaje natural. Su plataforma Dragon es ampliamente utilizada en aplicaciones de dictado y reconocimiento de voz en el sector salud. Nuance utiliza modelos avanzados de aprendizaje automático para ofrecer una alta precisión en la transcripción de voz y capacidades de síntesis de voz natural

otter.ai

Otter.ai ofrece servicios de transcripción automática que utilizan inteligencia artificial para convertir el habla en texto en tiempo real. Es especialmente útil en contextos de reuniones y conferencias, donde se necesita una transcripción rápida y precisa. Otter.ai soporta múltiples idiomas y ofrece características avanzadas como la identificación de altavoces y la exportación de transcripciones a diversos formatos

Enlaces de interés:

Charting the Course: Generative AI's Impact on Voice Synthesis in 2024

<https://blog.unrealspeech.com/exploring-2024s-generative-ai-in-voice-synthesis-technology/>

10 Key Predictions for the Future Of Voice Assistants

<https://masterofcode.com/blog/9-key-predictions-for-the-future-of-voice-assistants>

VoiceCloning – artículo super completo de sus pros y contras

<https://blog.marvik.ai/2023/03/21/state-of-the-art-in-voice-cloning-a-review/>

Parte 2

AWS Comprehend

Descripción:

AWS Comprehend es un servicio de procesamiento de lenguaje natural que utiliza aprendizaje automático para encontrar patrones y relaciones en datos de texto. Es capaz de realizar análisis de sentimientos, clasificación de texto, extracción de entidades...

Características:

- Análisis de Sentimientos: Determina el tono general de los textos.
- Extracción de Entidades: Identifica entidades clave como personas, lugares, organizaciones.
- Clasificación de Texto: Agrupa textos en categorías predefinidas.
- Análisis de Temas: Descubre temas en grandes colecciones de texto.

Tecnología:

Utiliza modelos de Deep learning entrenados en grandes conjuntos de datos.

Usos Comunes:

Análisis de opiniones en redes sociales, categorización de documentos, bussiness Intelligence...

Enlaces de interés:

En los productos de AWS realmente se necesita poco más que sus recursos y documentaciones, ya que son extremadamente completas y detalladas.

<https://aws.amazon.com/es/comprehend/>

<https://docs.aws.amazon.com/pdfs/comprehend/latest/dg/comprehend-dg.pdf>

Guía inicial de comprehend:

<https://www.youtube.com/watch?v=T63LDIRbwxw>

Cohere

Descripción:

Cohere ofrece una plataforma de procesamiento de lenguaje natural que facilita la construcción de aplicaciones NLP personalizadas mediante APIs fáciles de usar.

Características:

- Modelos de Lenguaje Preentrenados: Utiliza modelos avanzados que pueden ser ajustados para tareas específicas.
- Clasificación de Texto: Permite categorizar textos en diversas categorías.
- Generación de Texto: Genera texto coherente basado en prompts específicos.

Tecnología:

Basado en modelos de transformación, incluyendo variantes avanzadas de GPT.

Usos Comunes:

Generación de contenido, asistencia en redacción, análisis de texto.

Enlaces de interés:

<https://cohere.com/>

<https://docs.cohere.com/docs/quick-start-guides>

Video rápido de uso de cohere: https://www.youtube.com/watch?v=b_IKo6h-mX4

Deepset

Descripción:

Deepset proporciona herramientas para construir motores de búsqueda semánticos y sistemas de preguntas y respuestas, basados en modelos NLP avanzados.

Características:

- Haystack: Una plataforma que permite construir sistemas de búsqueda y QA (preguntas y respuestas) utilizando modelos de lenguaje.
- Modelos Personalizables: Posibilidad de ajustar modelos a necesidades específicas del dominio.

Tecnología:

Utiliza modelos BERT y otros modelos de lenguaje preentrenados.

Usos Comunes:

Sistemas de búsqueda avanzada, chatbots, asistentes virtuales.

Enlaces de interés:

<https://www.deepset.ai>

Deepset dispone de webinars así como de ebooks para aprender a usarlo:

<https://www.deepset.ai/webinars>

<https://www.deepset.ai/ebooks>

Tutorial de cómo usar su plataforma Haystack:

<https://www.youtube.com/watch?v=pOnkGdOvYfo>

Rasa

Descripción:

Rasa es una plataforma con una versión de código abierto para construir asistentes de conversación. Su versión pro está principalmente enfocada a empresas.

Características:

- Rasa Open Source: Motor de NLU (comprensión del lenguaje natural) y marco de diálogo para crear asistentes conversacionales.
- Rasa X: Herramienta para mejorar y gestionar asistentes conversacionales en producción.

Tecnología:

Basado en redes neuronales y técnicas de aprendizaje profundo.

Usos Comunes:

Chatbots para servicio al cliente, asistentes virtuales.

Enlaces de interés (Centrados en la versión opensource):

<https://rasa.com>

<https://rasa.com/docs/rasa/>

Tutorial de construcción de chatbot with rasa: (no lo he podido ver entero pero tiene buena pinta, lo mantengo para revisión futura)

<https://www.youtube.com/watch?v=B9QYBCiOxIE>

Lilt

Descripción:

Lilt ofrece una plataforma de traducción asistida por inteligencia artificial que mejora la eficiencia y precisión de la traducción humana.

Características:

- Traducción en Tiempo Real: Traducción y sugerencias en tiempo real para traductores humanos.
- Memoria de Traducción: Almacena traducciones previas para mejorar la consistencia y rapidez.

Tecnología:

Utiliza modelos de traducción neuronal.

Usos Comunes:

Traducción de contenido técnico, localización de software, documentación multilingüe.

Enlaces de interés:

<https://lilt.com>

<https://developers.lilt.com/quickstart>

del canal oficial de lilt <https://www.youtube.com/watch?v=YByEqFvzMw0>

Hugging Face

Descripción:

Hugging Face ofrece una amplia gama de herramientas y modelos preentrenados para NLP, disponibles a través de su biblioteca Transformers.

Características:

- Transformers Library: Acceso a modelos preentrenados para tareas como clasificación de texto, generación de texto, traducción, y más.
- Modelos Personalizables: Los modelos pueden ser afinados para tareas específicas del usuario.

Tecnología:

Utiliza una variedad de modelos de transformación avanzados, incluyendo BERT, GPT-3, T5, entre otros.

Usos Comunes:

Investigación en NLP, desarrollo de aplicaciones basadas en lenguaje natural, chatbots... Realmente tiene tal cantidad de herramientas que puede usarse para prácticamente cualquier cosa que te plantes en el mundo de la IA

Enlaces de interés:

<https://huggingface.co>

<https://huggingface.co/docs>

<https://huggingface.co/docs/chat-ui/index>

Disponen de un curso bastante bueno de NLP, algo que definitivamente estudiare más en profundidad a lo largo del curso

<https://huggingface.co/learn/nlp-course/chapter1/1>