



POR LA **ACCESIBILIDAD** EN LOS
MEDIOS **AUDIOVISUALES**

Tecnologías de transcripción en la nube y su aplicación al subtulado accesible



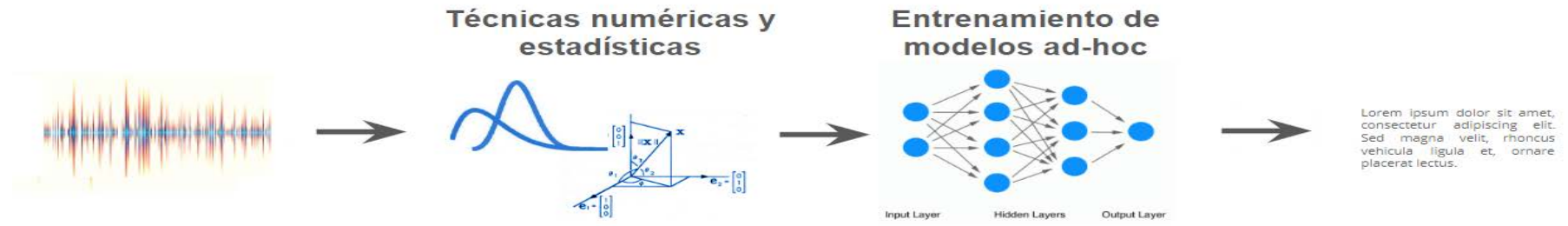
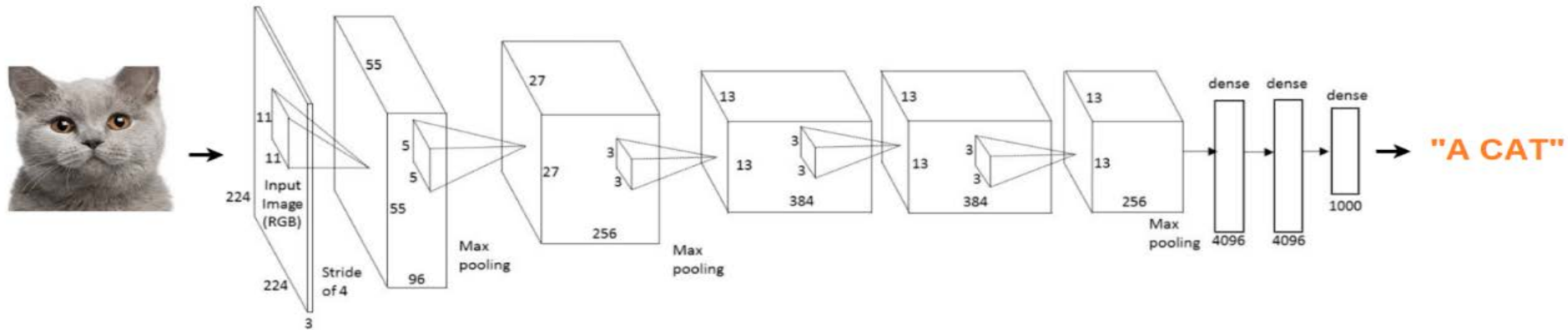
Introducción

- El subtitulado es el medio principal para acceder a los medios audiovisuales como televisión, cine, teatros para las personas con discapacidad auditiva.
- Para garantizar su calidad, en España el subtitulado accesible se rige por la norma UNE 153010 (AENOR)
- Cumplir estos criterios de calidad es un reto en el caso del subtitulado en tiempo real.
- En la mayoría de las televisiones, el subtitulado en tiempo real se realiza con la técnica del rehablado.



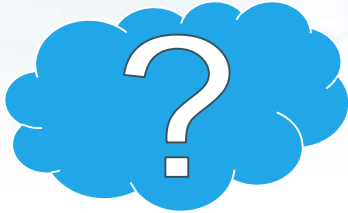
Marimar Rufino (Universidad de Montréal)
Rehablando en TVA (Montreal)

¿Por qué la Nube?



- Independencia del hablante
- Se ofrecen como servicio, no requieren infraestructura

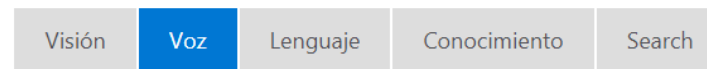
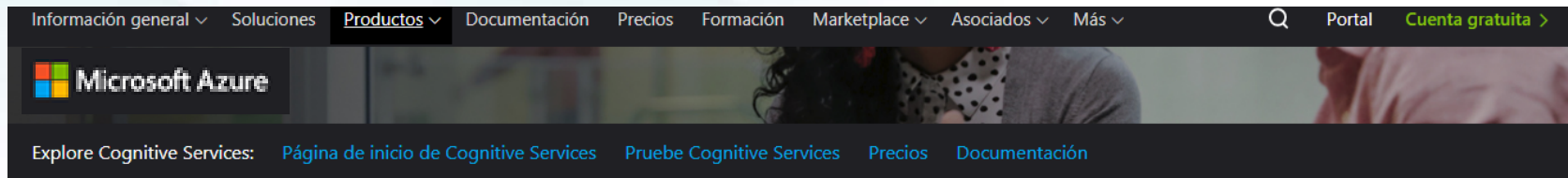
¿Por qué la Nube?




- Concentración de servidores en enormes CPDs
- Capacidad de computación y almacenamiento de datos ilimitadas
- Concentración de talento e imaginación
- Espacio de desarrollo, innovación y mejoras continuas.



Tecnologías de Transcripción en la Nube



Descubra cómo Voz permite la integración de capacidades de procesamiento de voz en cualquier aplicación o servicio. Convierta el lenguaje hablado en texto o produzca una voz natural a partir de texto mediante fuentes de voz estándares (o personalizables). Pruebe cualquier servicio gratis y compile rápidamente aplicaciones y servicios basados en voz con las siguientes capacidades.




Conversión de voz en texto

[Pruébelo gratis](#) | [Más información](#)

Reconocimiento de voz y transcripción de voz automáticos (voz a texto) | [Demo](#)

Reconocimiento de voz y transcripción de voz personalizables (voz a texto) | [Demo](#)




Speaker Recognition VERSIÓN PRELIMINAR

[Pruébelo gratis](#) | [Más información](#)

Identificación del hablante | [Demo](#)

Comprobación del hablante | [Demo](#)




Text to Speech

[Pruébelo gratis](#) | [Más información](#)

Texto a voz automático | [Demo](#)

Fuentes de voz personalizables para texto a voz | [Demo](#)



Traducción de voz

[Pruébelo gratis](#) | [Más información](#)

Traducción en tiempo real

Traducción de voz automatizada

Traducción personalizable

Tecnologías de Transcripción en la Nube

PRODUCTOS Y SERVICIOS

Protege tus datos y hardware, obtén información valiosa en tiempo real y aumenta tu productividad con las mismas herramientas abiertas e inteligentes que utiliza Google.

[VER LA DOCUMENTACIÓN PARA DESARROLLADORES](#)

[VER MI CONSOLA](#)



Compute >

Compute Engine >

Despliega máquinas virtuales en la infraestructura de Google

App Engine >

Plataforma como servicio (PaaS) para aplicaciones y backends

Kubernetes Engine >

Despliega contenedores en Google Cloud Platform

Cloud Functions ^{BETA} >

Entorno sin servidor para crear y conectar servicios en la nube



Big Data >

RioQuerv



Almacenamiento y bases de datos >

Cloud Storage >

Almacenamiento de objetos en caché perimetral global

Cloud SQL >

Servicio de bases de datos MySQL y PostgreSQL totalmente administrado

Cloud Bigtable >

Servicio de base de datos NoSQL totalmente administrado

Cloud Spanner >

Servicio de base de datos relacional para aplicaciones esenciales

Cloud Datastore >

Base de datos NoSQL para datos no relacionales

Disco persistente



Cloud AI >

AutoML de Cloud ^{ALFA} >

Prepara modelos de aprendizaje automático personalizados y de alta calidad sin apenas esfuerzo y sin necesidad de conocimientos avanzados sobre este ámbito

TPU de Cloud ^{BETA} >

Accede a los mismos aceleradores que utiliza Google para desarrollar productos de aprendizaje automático de primera categoría

Cloud Machine Learning Engine >

Aprendizaje automático sin restricciones de datos ni de tamaño

Cloud Job Discovery ^{BETA PRIVADA} >

Potencia las capacidades de tu sitio de ofertas de empleo con el aprendizaje automático

Dialogflow Enterprise Edition ^{BETA} >

Entabla conversaciones en diferentes dispositivos y plataformas

Cloud Natural Language >

Análisis de sentimiento a través de texto no estructurado

API Speech de Cloud >

Conversión de voz en texto con la tecnología de aprendizaje automático

Síntesis de voz de Cloud ^{BETA} >

Conversión de texto en voz con la tecnología del aprendizaje automático

API Cloud Translation >

Traduce de forma dinámica entre la mayoría de pares de idiomas

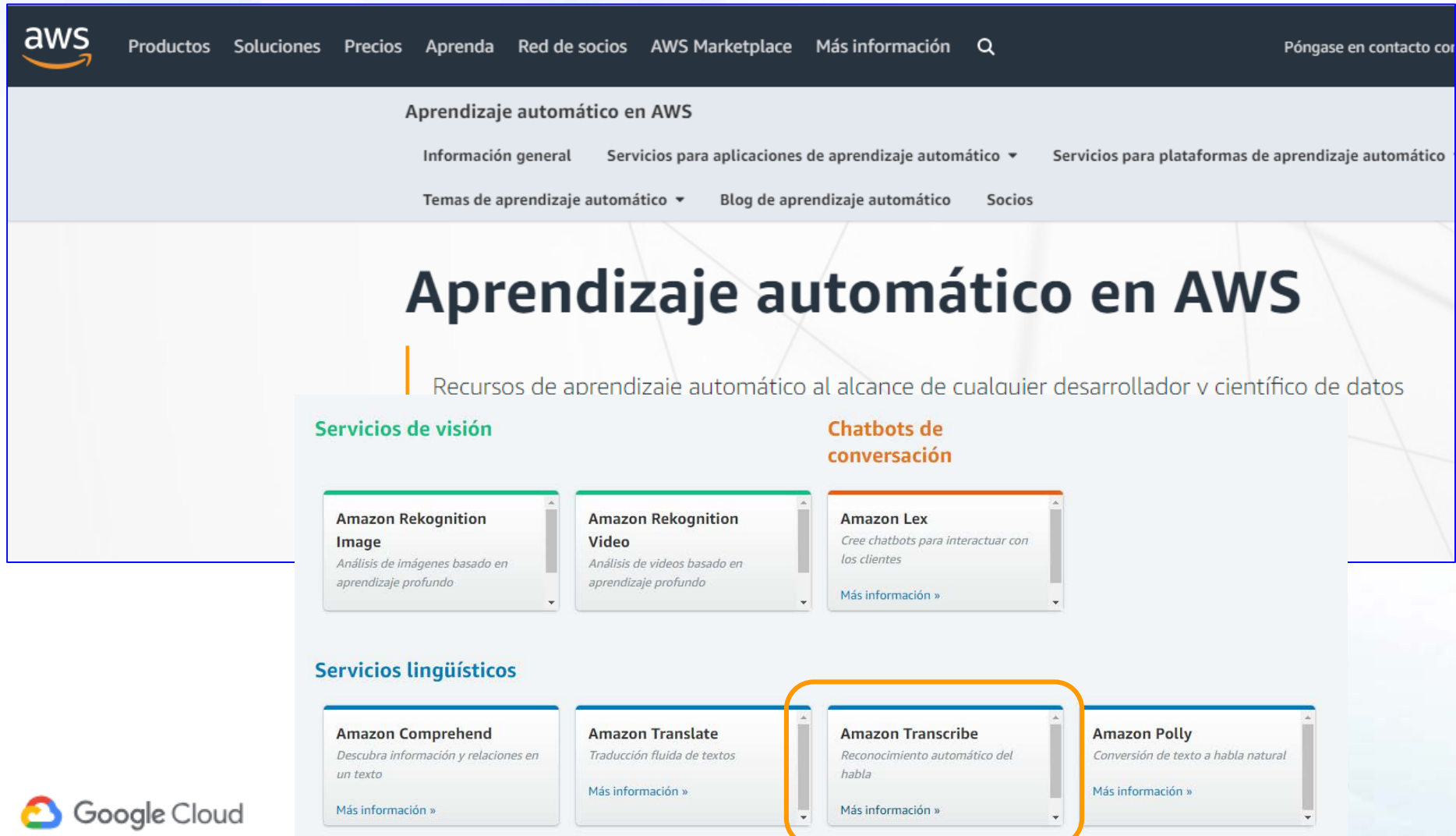
API Vision de Cloud >

Consigue información valiosa a partir de imágenes gracias al aprendizaje automático

API Video Intelligence de Cloud >

Extrae metadatos de vídeos

Tecnologías de Transcripción en la Nube



aws Productos Soluciones Precios Aprenda Red de socios AWS Marketplace Más información Póngase en contacto con

Aprendizaje automático en AWS

Información general Servicios para aplicaciones de aprendizaje automático Servicios para plataformas de aprendizaje automático

Temas de aprendizaje automático Blog de aprendizaje automático Socios

Aprendizaje automático en AWS

Recursos de aprendizaje automático al alcance de cualquier desarrollador y científico de datos

Servicios de visión

Amazon Rekognition Image

Análisis de imágenes basado en aprendizaje profundo

Amazon Rekognition Video

Análisis de videos basado en aprendizaje profundo

Chatbots de conversación

Amazon Lex

Cree chatbots para interactuar con los clientes

[Más información »](#)

Servicios lingüísticos

Amazon Comprehend

Descubra información y relaciones en un texto

[Más información »](#)

Amazon Translate

Traducción fluida de textos

[Más información »](#)

Amazon Transcribe

Reconocimiento automático del habla

[Más información »](#)

Amazon Polly

Conversión de texto a habla natural

[Más información »](#)



Evaluación

Evaluamos sobre una misma muestra la calidad de **Rehablado** versus la calidad de **Transcripción con Google Speech API**

Dimensiones Evaluadas

Precisión	<p>Medida de la fidelidad de los subtítulos respecto al discurso original.</p> <ul style="list-style-type: none"> • Cálculo basado en el modelo NER, NER >= 98% (UNE 153010)
Literalidad	<p>Los subtítulos deben ser lo más próximos al original.</p>
Latencia	<p>Tiempo transcurrido desde que la palabra aparece en el discurso hasta que aparece en el subtítulo.</p> <ul style="list-style-type: none"> • < 8 seconds (UNE 153010)
Velocidad	<p>Número máximo de caracteres por segundo (cps).</p> <ul style="list-style-type: none"> • No debe superar los 15 cps. (UNE 153010)

Muestra



Modelo NER

$$precision = \frac{N - E - R}{N} \times 100$$

N = Número total de palabras en los subtítulos (incluyendo signos de puntuación, cambio de hablante, etc.)

E= Número total de errores de edición (omisiones, adiciones incorrectas, uso incorrecto de la puntuación, etc.)

R= Número total de errores de reconocimiento

Tipo de Error	Puntuación	Descripción	Ejemplo (entre paréntesis las palabras originales u omitidas)
grave	1	cambia el significado original, el error puede no ser evidente	<i>En el interior hay problemas, (pero no tantos como en la costa). Ahí sí que se necesita ayuda.</i>
normal	0,5	el error causa extrañeza (omisiones, sustituciones)	<i>Los detalles que nadie de (ve) son importantes</i>
leve	0,25	no dificultan la comprensión (mayúsculas incorrectas, pequeñas omisiones)	<i>Tan poco (tampoco) lo tiene</i>

Evaluación

Resultados

Subtitulado	Rehablado	Google Speech
Cálculo NER	98,6%	95,6%
Cálculo NER (sin incluir errores de puntuación, identificación de hablante)	98,8%	98,1%
Latencia media	6-7 segundos en las secciones en directo	1-2 segundos
Velocidad media	10-12 cps	13-14 cps
Precisión/Literalidad	Muchas omisiones con cambios rápidos de interlocutor, o ambientes ruidosos, pero capacidad de sintetizar la situación Capacidad de inferir palabras mal pronunciadas.	Muy buena literalidad con un único interlocutor Muchas omisiones con cambios rápidos de interlocutor, o ambientes ruidosos No infiere palabras mal pronunciadas.

Evaluación

Subtitulado	Rehablado	Google Speech
Cálculo NER	98,6%	95,6%
Cálculo NER (sin incluir errores de puntuación, identificación de hablante)	98,8%	98,1%
Latencia media	6-7 segundos en las secciones en directo	1-2 segundos



Conclusiones

- Las tecnologías de transcripción en la nube no alcanzan el nivel de precisión requerido para un subtítulo de calidad en directo, pero se están acercando.
- Tienen la literalidad y el nivel de precisión necesarios para plantear un cambio en la técnica del subtítulo en directo en los canales de TV.
- El rehablado podría eliminarse, y mantenerse la función de revisión del texto generado automáticamente.
- Con las progresivas mejoras en la transcripción, como puede ser incorporación de puntuación, o la identificación de hablantes, la técnica seguiría siendo válida aunque las correcciones necesarias irían disminuyendo.
- Consideramos pues que la aplicación de estas tecnologías es inmediata y pueden suponer una mejora considerable para simplificar el subtítulo en tiempo real, resolviendo los problemas de sincronización, dependencia del locutor, y pérdida de literalidad en los casos de hablantes muy rápidos.