# *IA Conversacional: más allá del "Perdona, no te he entendido"*

**Nieves Ábalos (@nieves_as)**

Founder & Chief Product Officer, Monoceros Labs

# ¡Hola!

**Nieves Ábalos Serrano**

Chief Product Officer &
Cofundadora de Monoceros Labs

Twitter: @nieves_as

LinkedIn: /in/nievesabalosserrano

- **UGR**: Ingeniera Software, Máster y PhD sin terminar en sistemas de diálogo.
- **Dpto Innovación** BEEVA (BBVA Next Technologies): I+D en conversacional y NLP.
- **Monoceros Labs:** emprendedora, estrategia y gestión de producto, diseño conversacional y desarrollo conversacional (lo que toque).
- **Alexa Champion.**
- **Women in Voice Spain (Lead)**
- **Comunidad de Alexa en español.**

# Perdona, no te he entendido

# IA conversacional

# IA Conversacional

amazon alexa

Mis consolas Alexa | Identificarse

Alexa Skills Kit | Alexa Voice Service | Dispositivos conectados | Programas de Alexa | Docs | Blog de Alexa

## Conversational AI: Computers That Talk

Conversational AI systems are computers that people can interact with simply by having a conversation, our most natural form of interaction. In short, it is what allows us to talk to voice-driven technologies like Amazon Alexa and ask about the weather, order products online, and even call a cab, simply by using the language we already know.

With conversational AI, voice-enabled devices like Amazon Echo are finally starting to enable the sort of magical interactions we've dreamed of for decades (think: Star Trek computer). Through a voice user interface (VUI), voice services like Alexa can communicate with people in ways that feel effortless, solve problems, and get smarter over time.

## Teaching Computers to Converse

Conversations can be conceptually and emotionally complex; they entail much more than simple input and output. When we talk to each other, how we say something also matters as much as ***what*** we say.

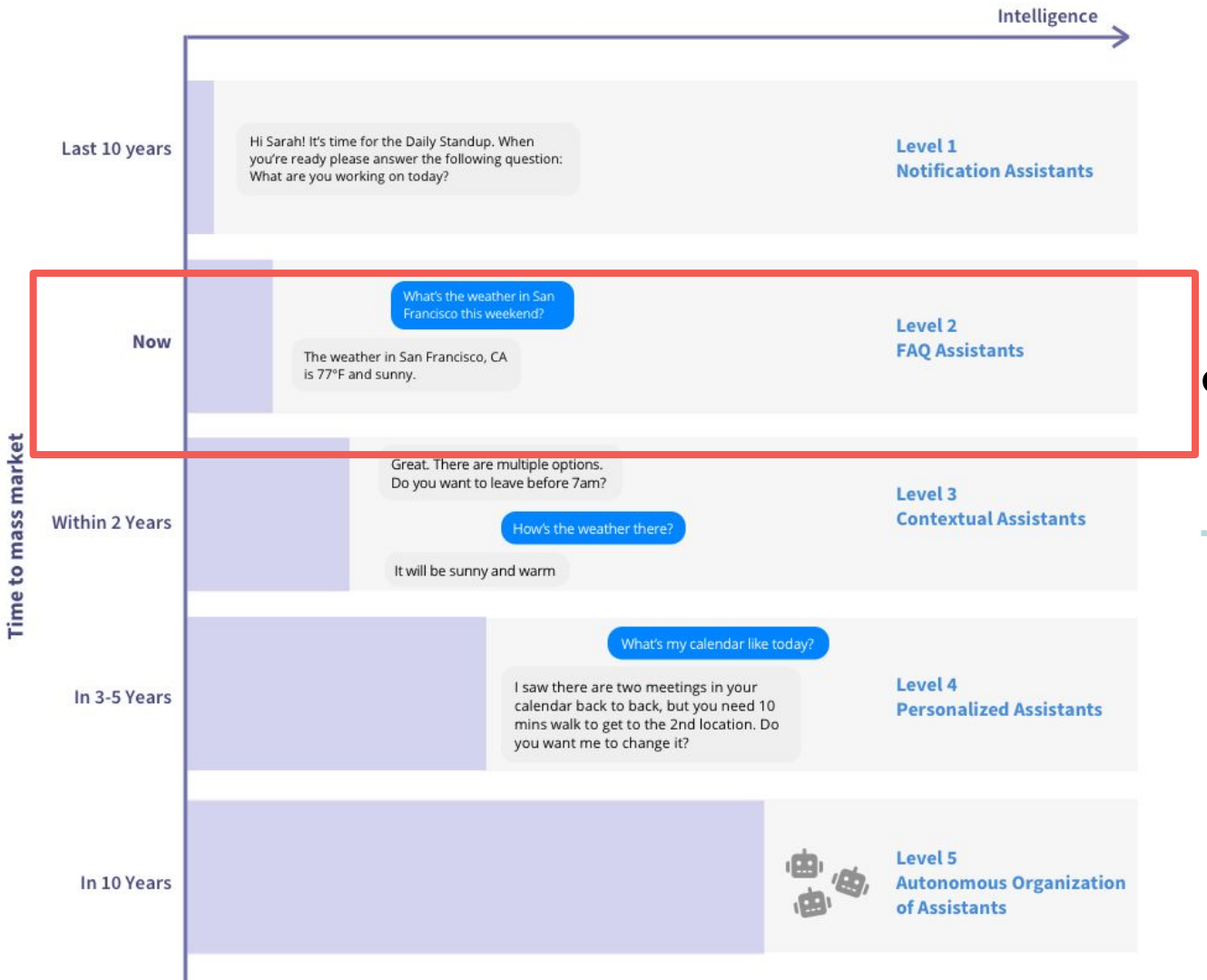
Computers can't grasp these nuances, and that's where voice design comes in. A well-designed VUI is flexible, and takes into account these unwritten rules of conversation. It enables computers to think and talk as we do, and not like a robot.

El objetivo es hacer que las personas interactuemos de **manera natural conversando con la tecnología**, teniendo en cuenta que las conversaciones pueden ser emocionalmente complejas, además de las reglas no escritas de la comunicación.

# Inteligencia Artificial + Canal / Modo


monoceros LABS

# Interfaces conversacionales hoy




monoceros LABS

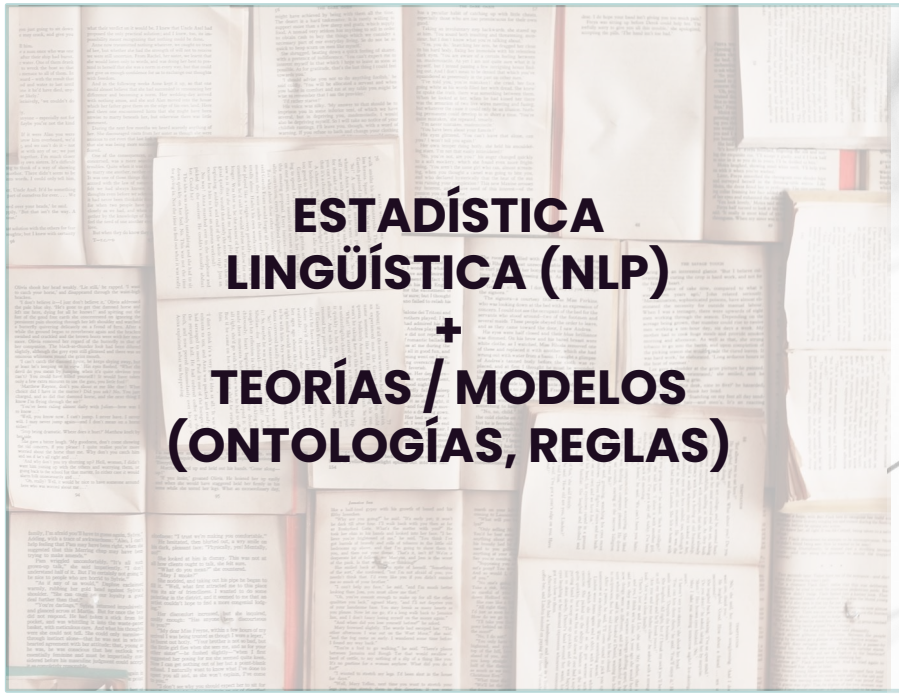
# ¿Dónde estamos entonces?

## Inteligencia vs Tiempo hasta llegar al mercado

Source: Alan Nichols, August 2018, O'Reilly

# Tecnologías: ayer y hoy

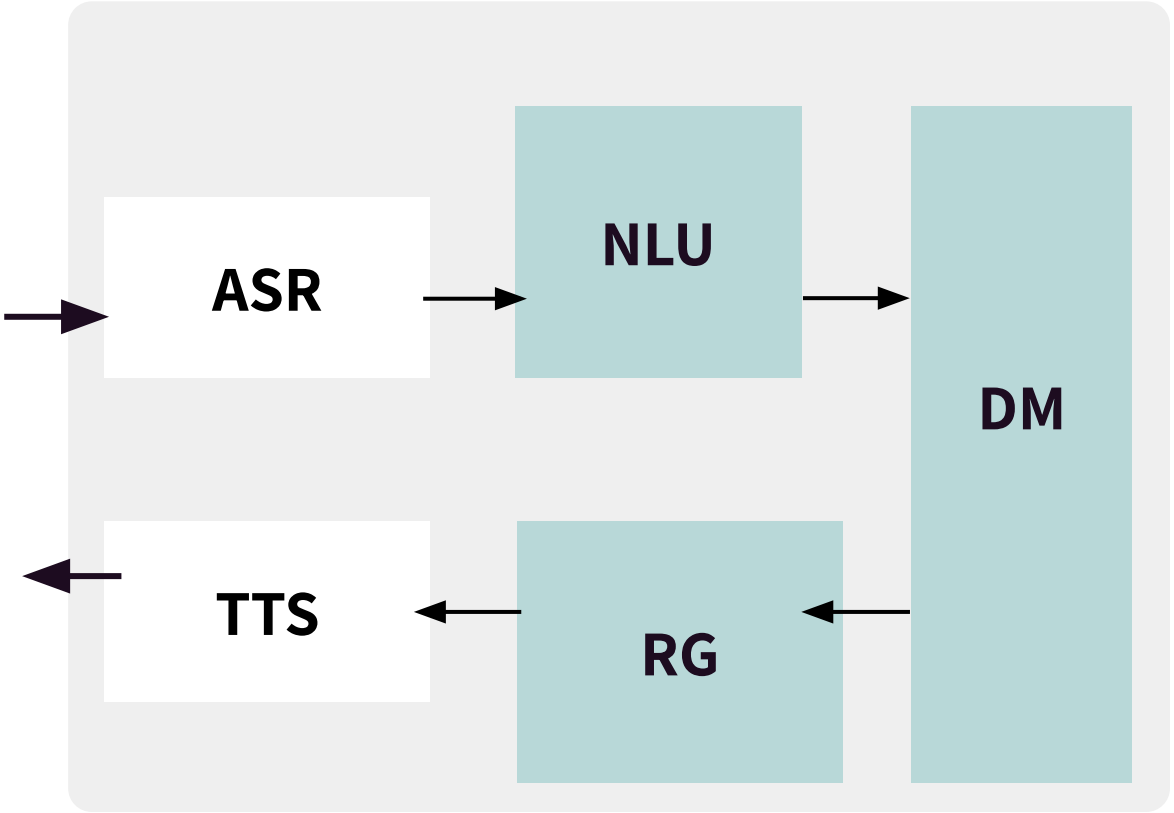
monoceros LABS

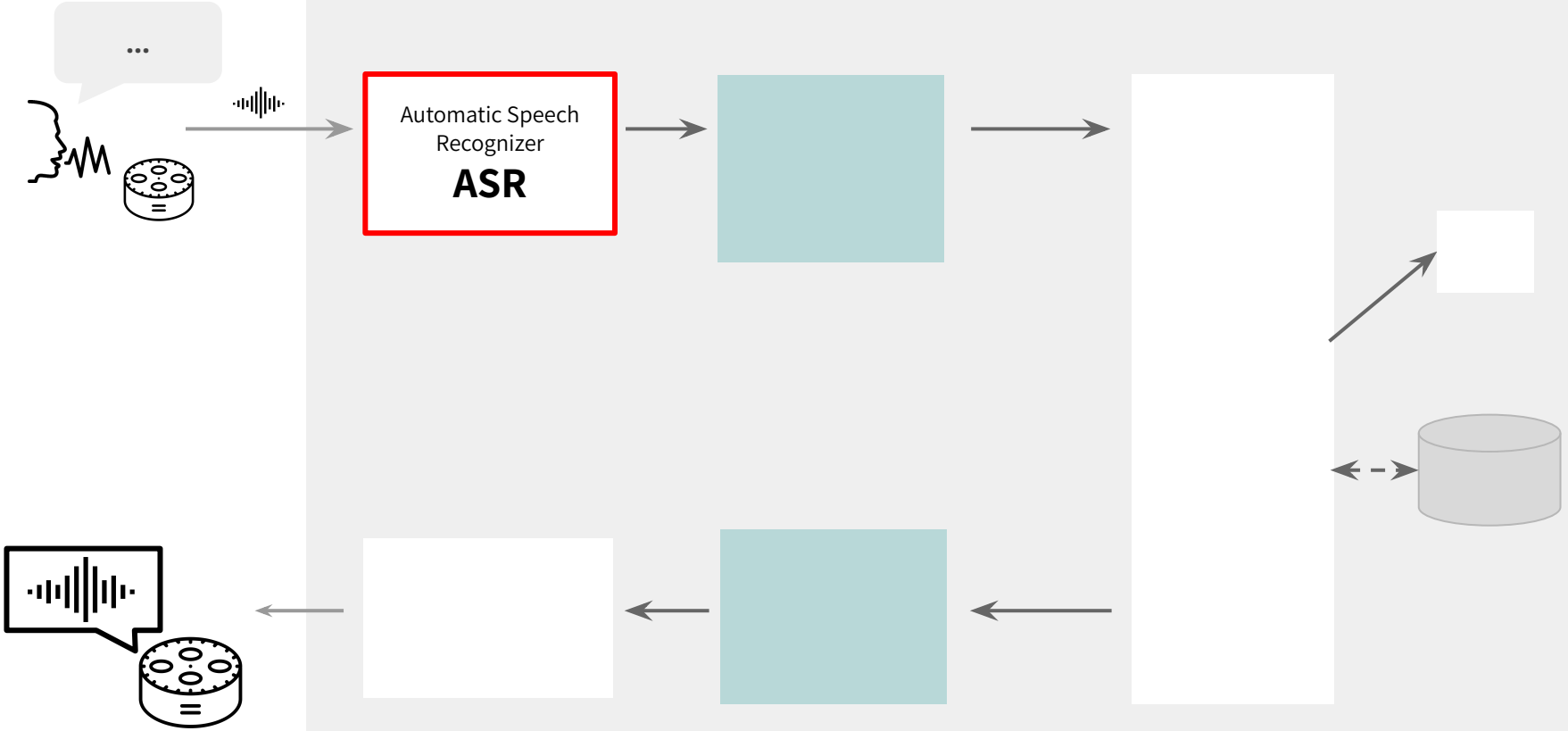



# ¿Es “inteligente” un asistente de voz?

monoceros LABS

“Quiero viajar a México?”



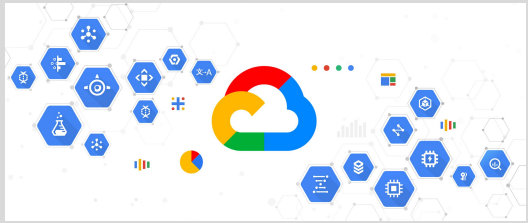

...
Automatic Speech Recognizer
ASR

# 2013… ¡ASR en el umbral de la capacidad humana!

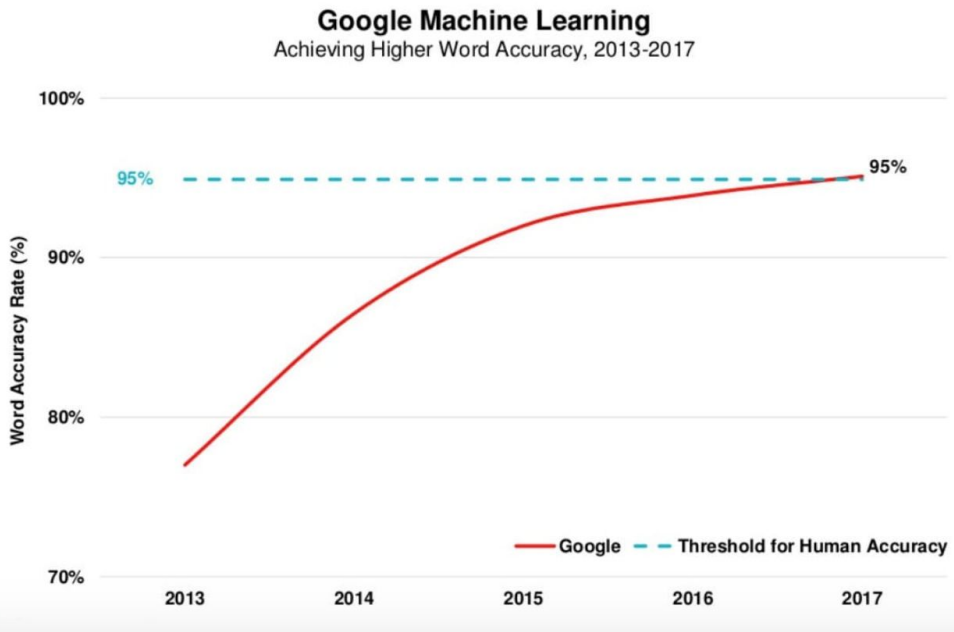
monoceros LABS

Google

La mejora en la capacidad de cómputo (nube) **redujo el tiempo** de procesamiento de **enormes cantidades de datos** (speech) >> mejores modelos de ASR



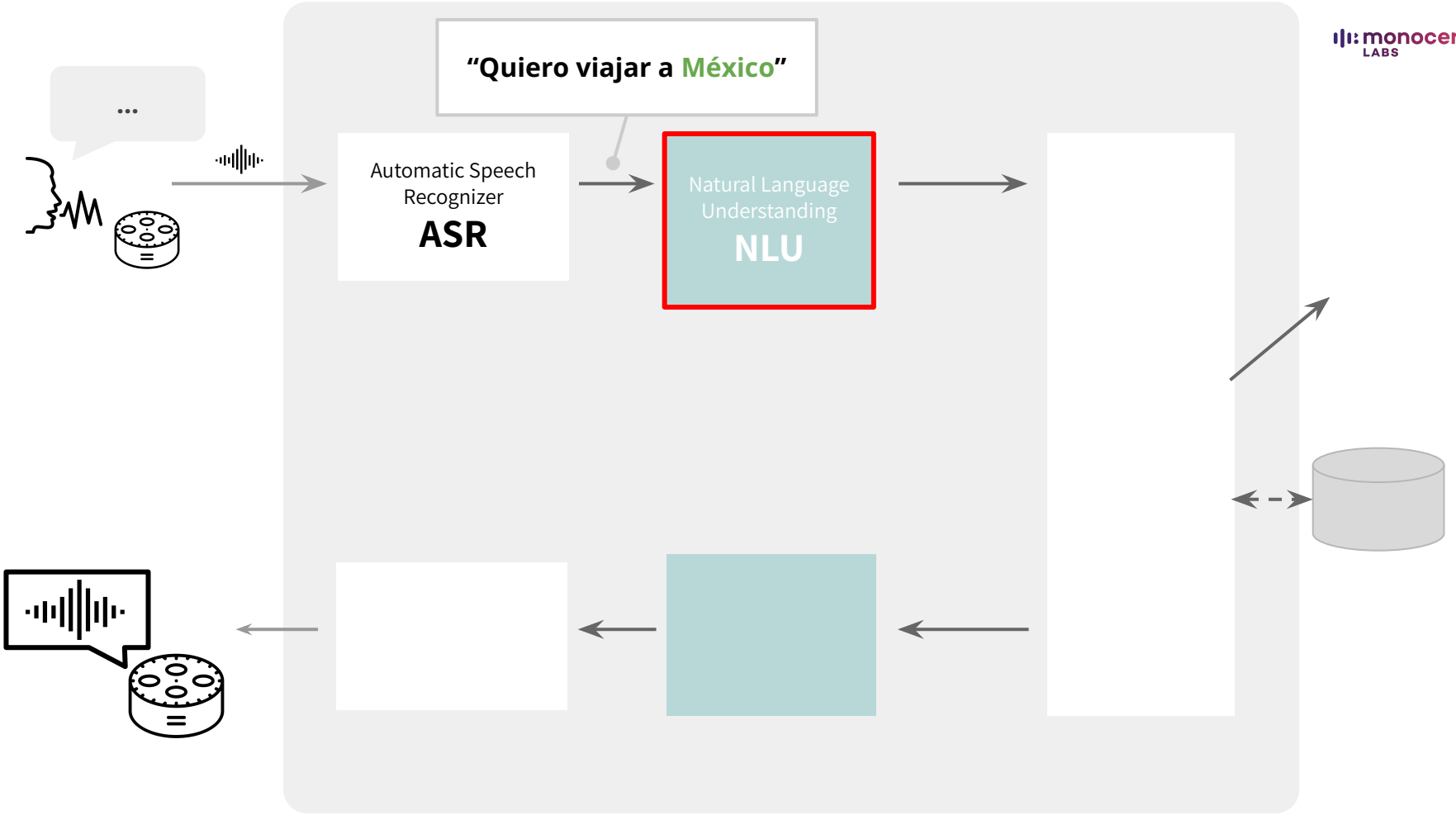
Computing power and artificial intelligence are largely behind the advances in this space. With massive amounts of speech data combined with faster processing, speech recognition has hit an inflection point where its capabilities are roughly on par with humans. The graph below is from Mary Meeker's 2017 Internet Trends report. It plots Google's word accuracy rate which recently broke the 95% threshold for human accuracy.



# El **ruido** y

*The Cocktail Party Problem*

El reto del ASR: tratar de identificar al hablante y separar lo que dice, en entornos de mucho ruido (como una fiesta).


monoceros
LABS
“Quiero viajar a México”
...
Automatic Speech
Recognizer
ASR
Natural Language
Understanding
NLU

# Arquitectura: NLU (Entrenando el modelo)

frase o *utterance*:

Quiero viajar a **México**

valor

entidad o *slot*
**lugar**

**viajarLugar**

intención del usuario o *intent*

- Viajo a **Estados Unidos**
- Quiero viajar a **Madrid**
- Viajaré a **Asia**
- Pues quiero un viaje a **Nueva York**
- ...

Clasificamos diferentes frases según la intención del usuario:

- aportamos **semántica**,
- indicamos la información relevante (entidades).

# Arquitectura: NLU (Entrenando el modelo)

frase o *utterance*:

Quiero viajar a {**México**}

valor

entidad o *slot* **ciudad**

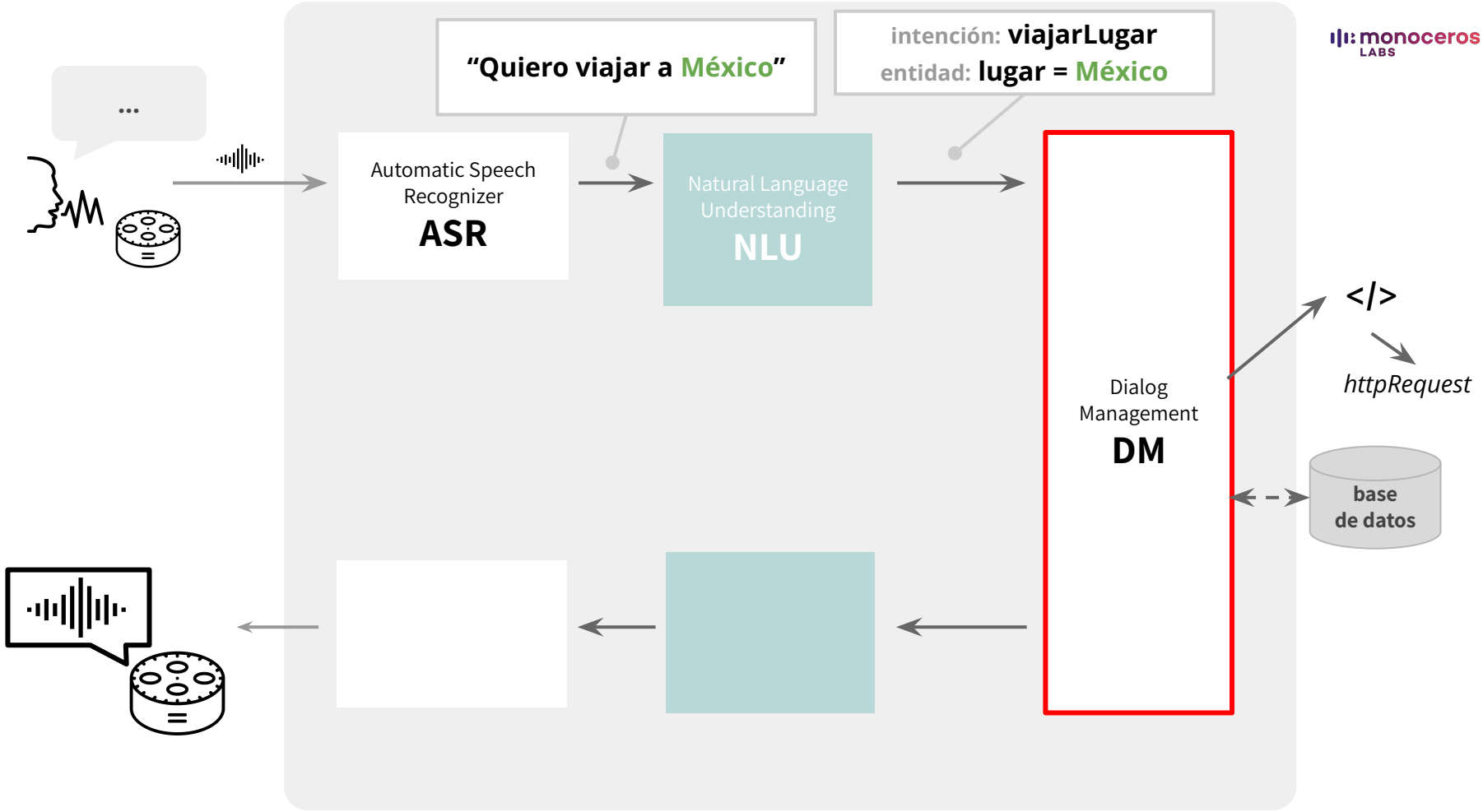
**viajarCiudad**

intención del usuario o *intent*

- Viajo a **Newark**
- Quiero viajar a **Madrid**
- Viajaré a **Bangkok**
- Pues quiero un viaje a **Nueva York**
- ...

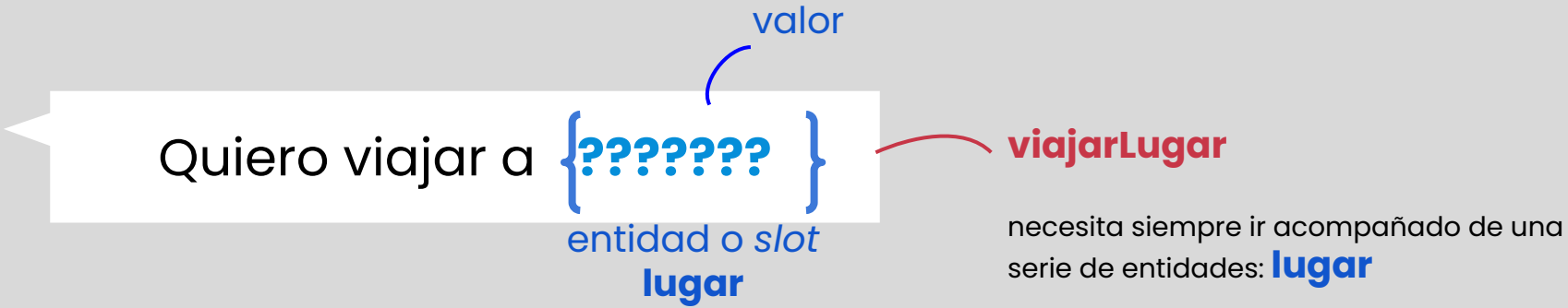
Clasificamos diferentes frases según la intención del usuario:

- aportamos **semántica**,
- indicamos la información relevante (entidades).


monoceros LABS
...
"Quiero viajar a México"
intención: viajarLugar
entidad: lugar = México
Automatic Speech Recognizer
ASR
Natural Language Understanding
NLU
Dialog Management
DM
</>
httpRequest
base de datos

# Arquitectura: DM > Gestión de errores

- **Slot filling**




# Arquitectura: DM > Gestión de errores

- **Slot filling**


**viajarLugar**

necesita siempre ir acompañado de una serie de entidades: **lugar**

**viajarLugar** | **lugar = ??** → ¡KO! → **ACCIÓN: Preguntar al Usuario por un LUGAR**

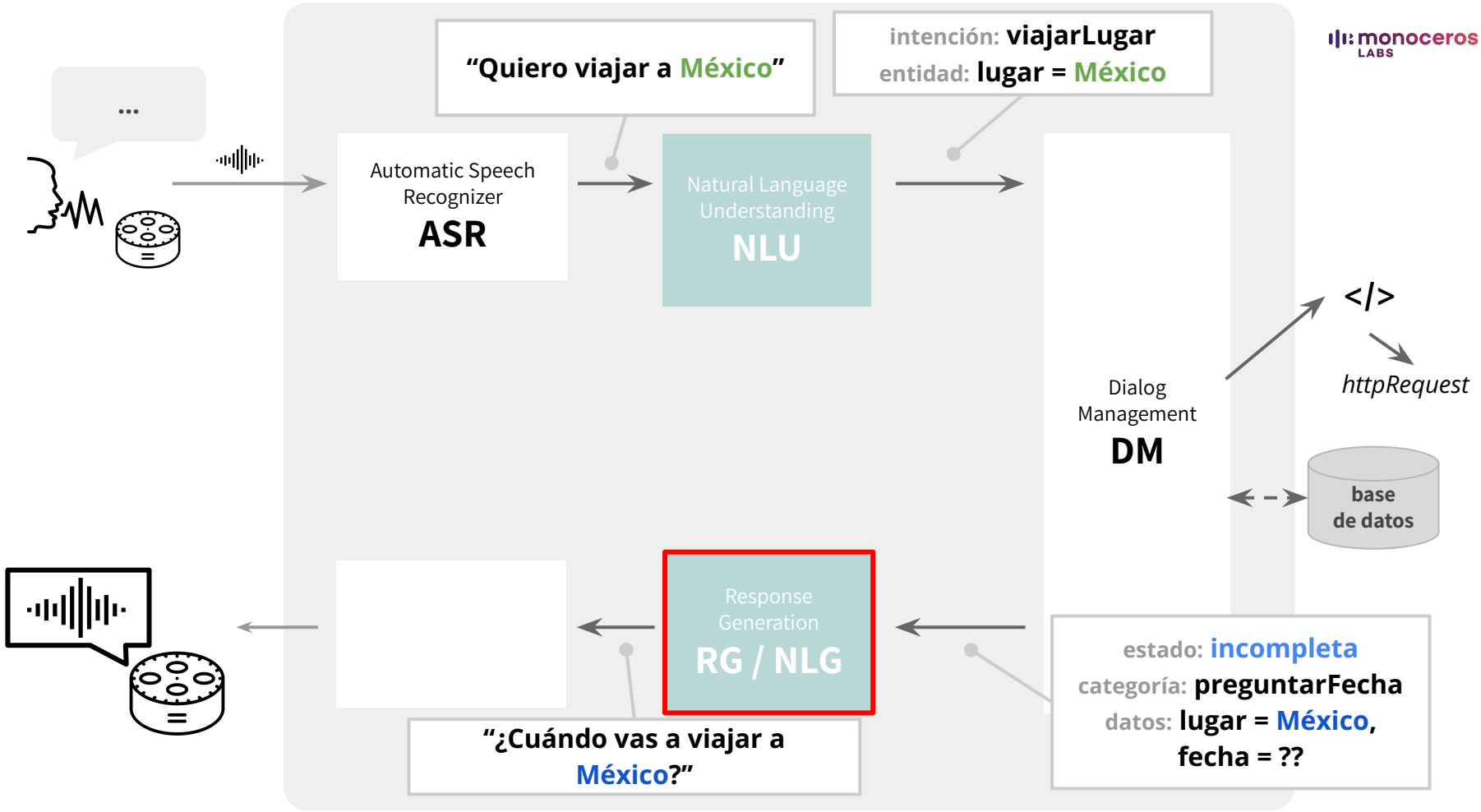
**viajarLugar** | **lugar = México** → ¡OK! → **ACCIÓN: Voy a consultar en nuestro sistema fechas disponibles para LUGAR, después pregunta por FECHA**

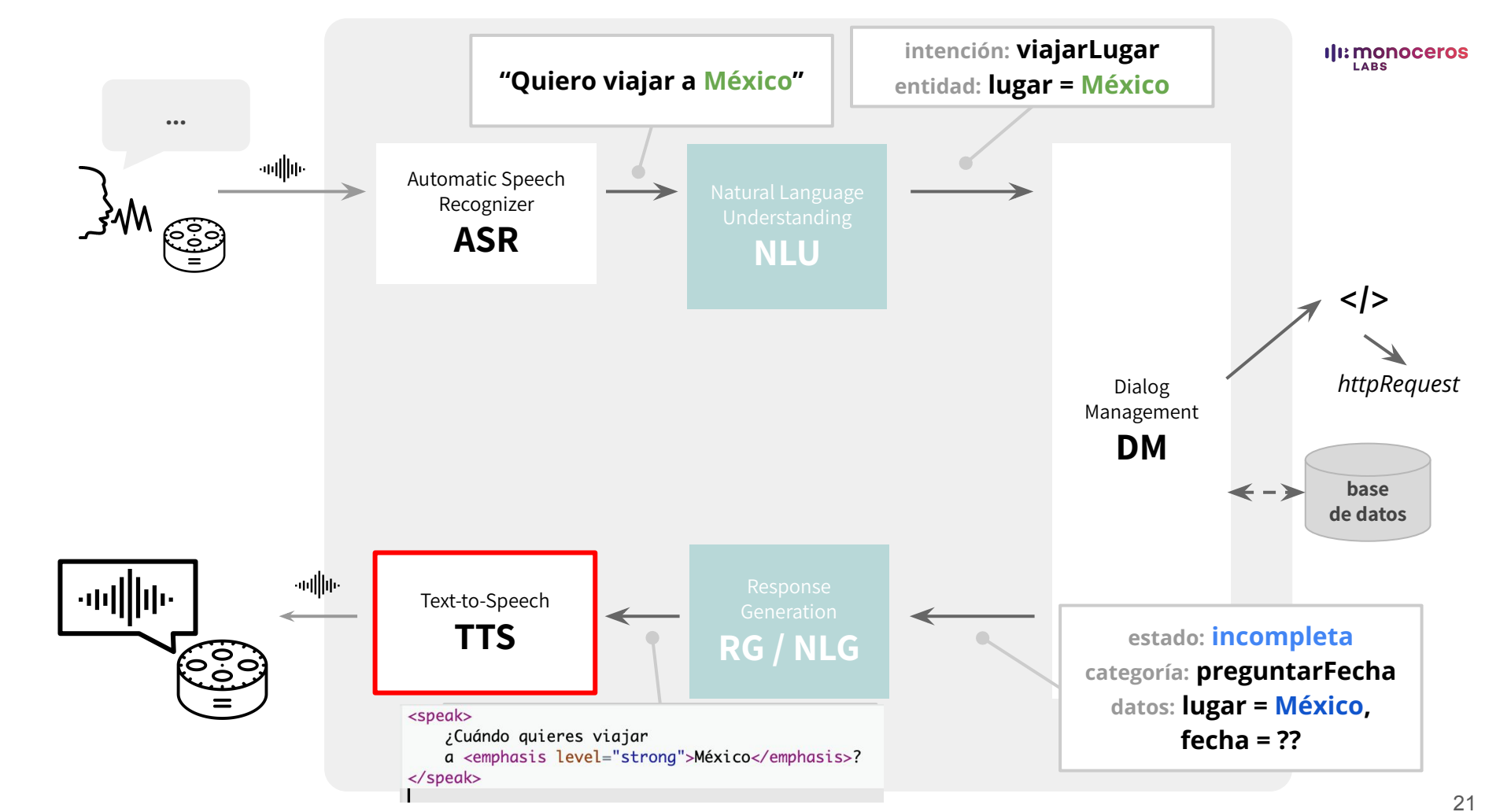
**fecha = ...**

**f(x){...}**
**API -> GET FECHAS**

DB

Nieves,
{"lugar": "México"}

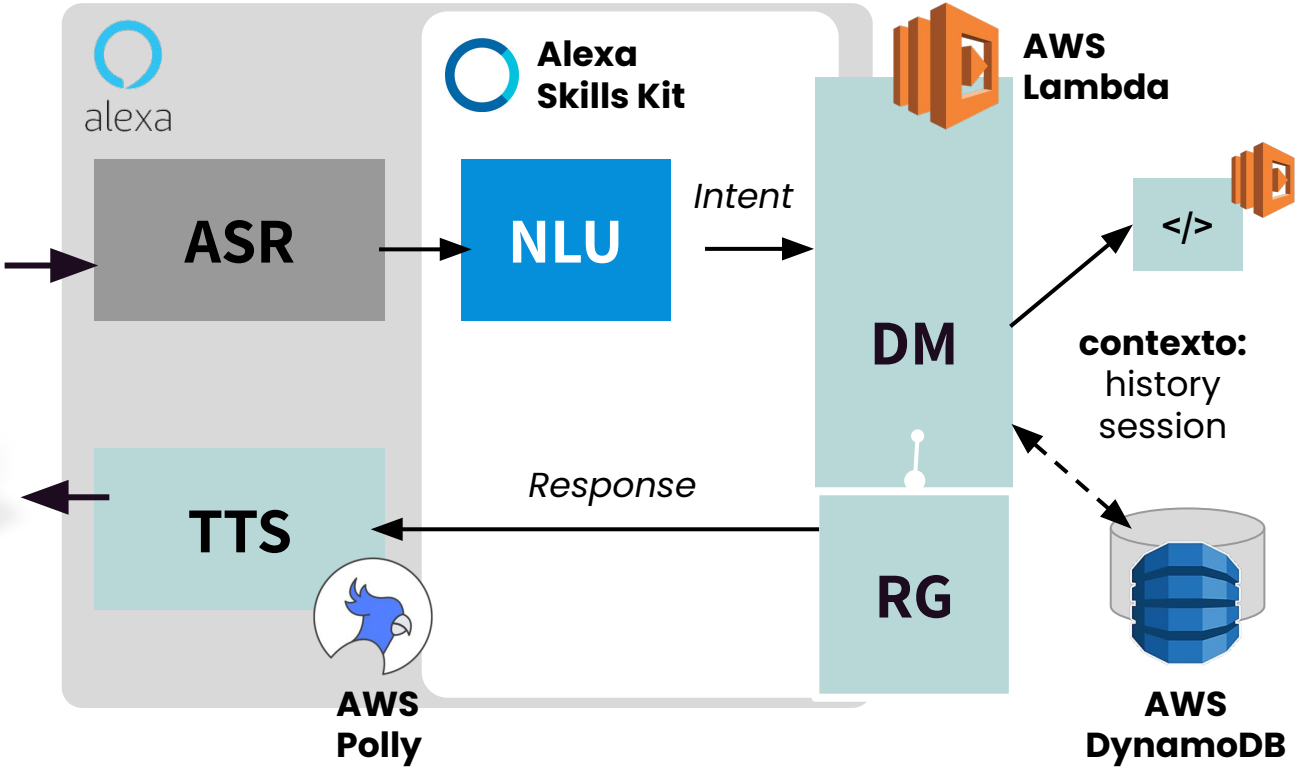

monoceros LABS
...
"Quiero viajar a México"
intención: viajarLugar
entidad: lugar = México
Automatic Speech Recognizer
ASR
Natural Language Understanding
NLU
Dialog Management
DM
</>
httpRequest
base de datos
Response Generation
RG / NLG
estado: incompleta
categoría: preguntarFecha
datos: lugar = México, fecha = ??
"¿Cuándo vas a viajar a México?"


...
"Quiero viajar a México"
intención: viajarLugar
entidad: lugar = México
monoceros LABS
Automatic Speech Recognizer
ASR
Natural Language Understanding
NLU
</>
httpRequest
Dialog Management
DM
base de datos
Text-to-Speech
TTS
Response Generation
RG / NLG
estado: incompleta
categoría: preguntarFecha
datos: lugar = México, fecha = ??
<speak>
¿Cuándo quieres viajar
a <emphasis level="strong">México</emphasis>?
</speak>

# Tecnologías subyacentes (Amazon Alexa)

monoceros LABS

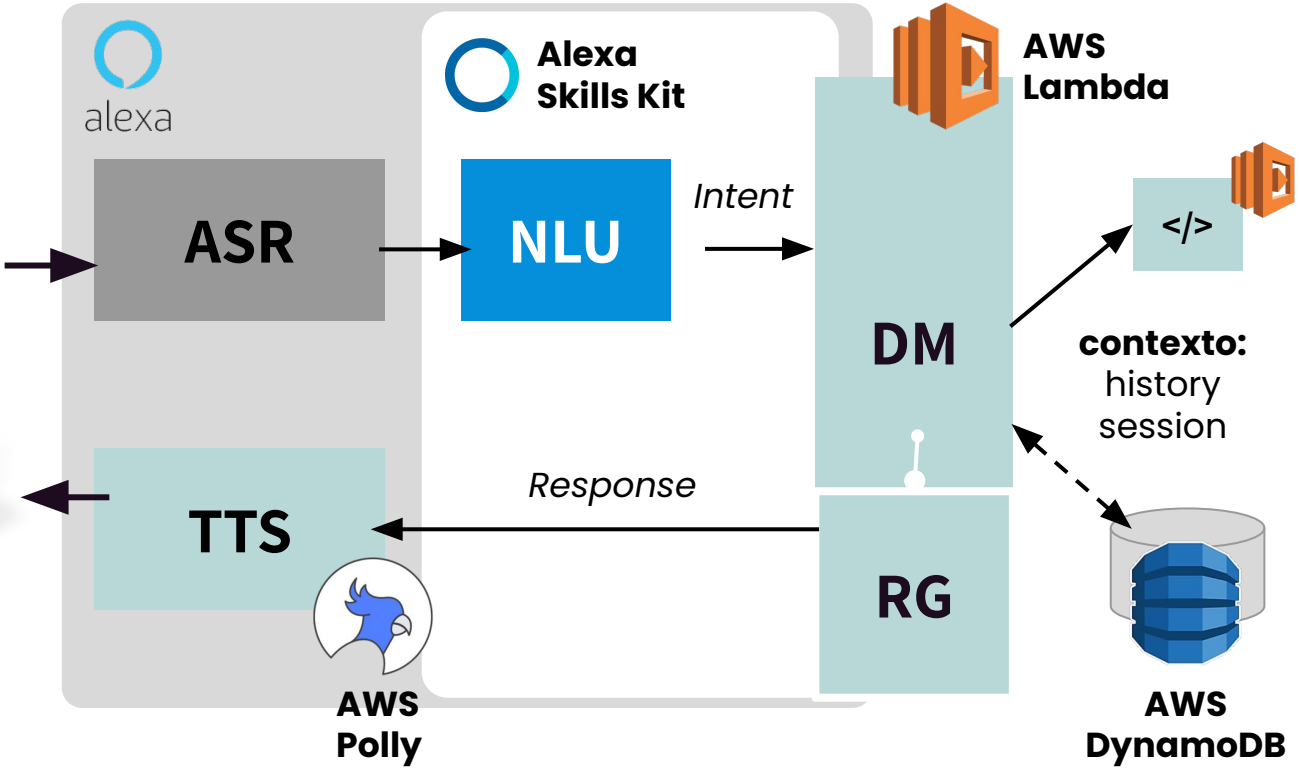
**Digo:**
"Alexa,
abre Veo Veo"



# Tecnologías subyacentes (Amazon Alexa)

monoceros LABS

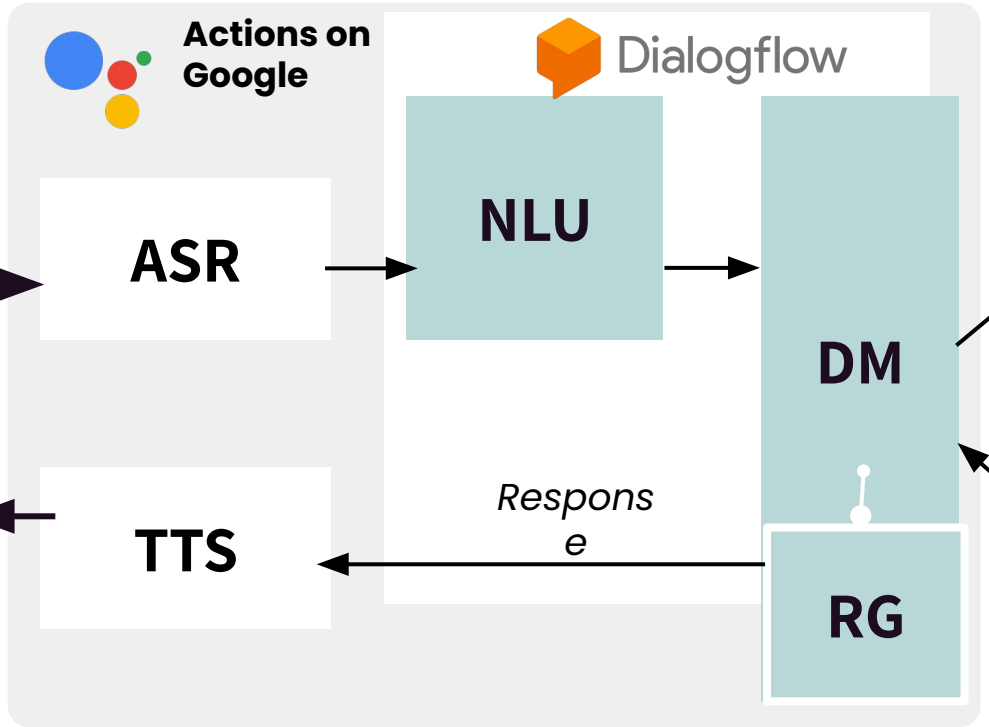
**Digo:**
"Alexa,
abre Veo Veo"



# Tecnologías subyacentes (Google Assistant)

monoceros LABS

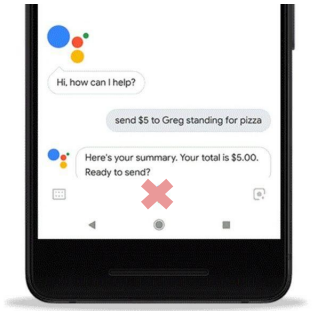
**Digo:**
"OK Google, abre Juego de Tronos"

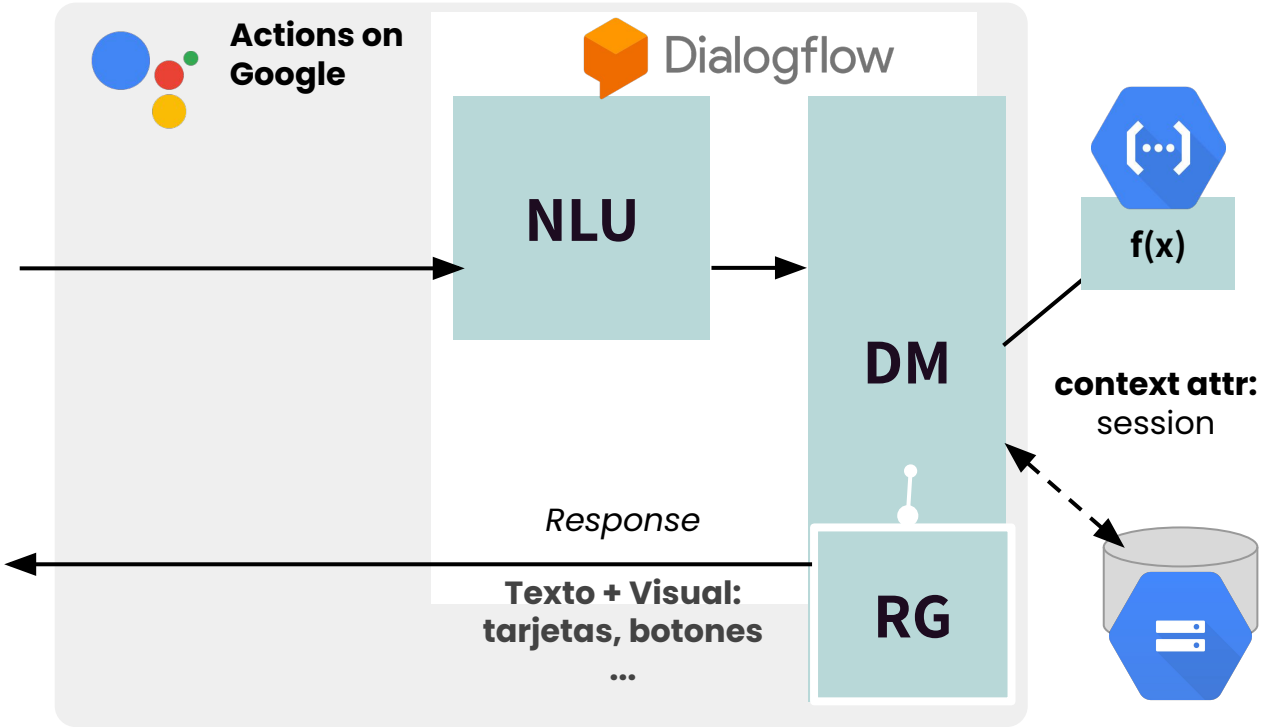


# Tecnologías subyacentes (Google Assistant)

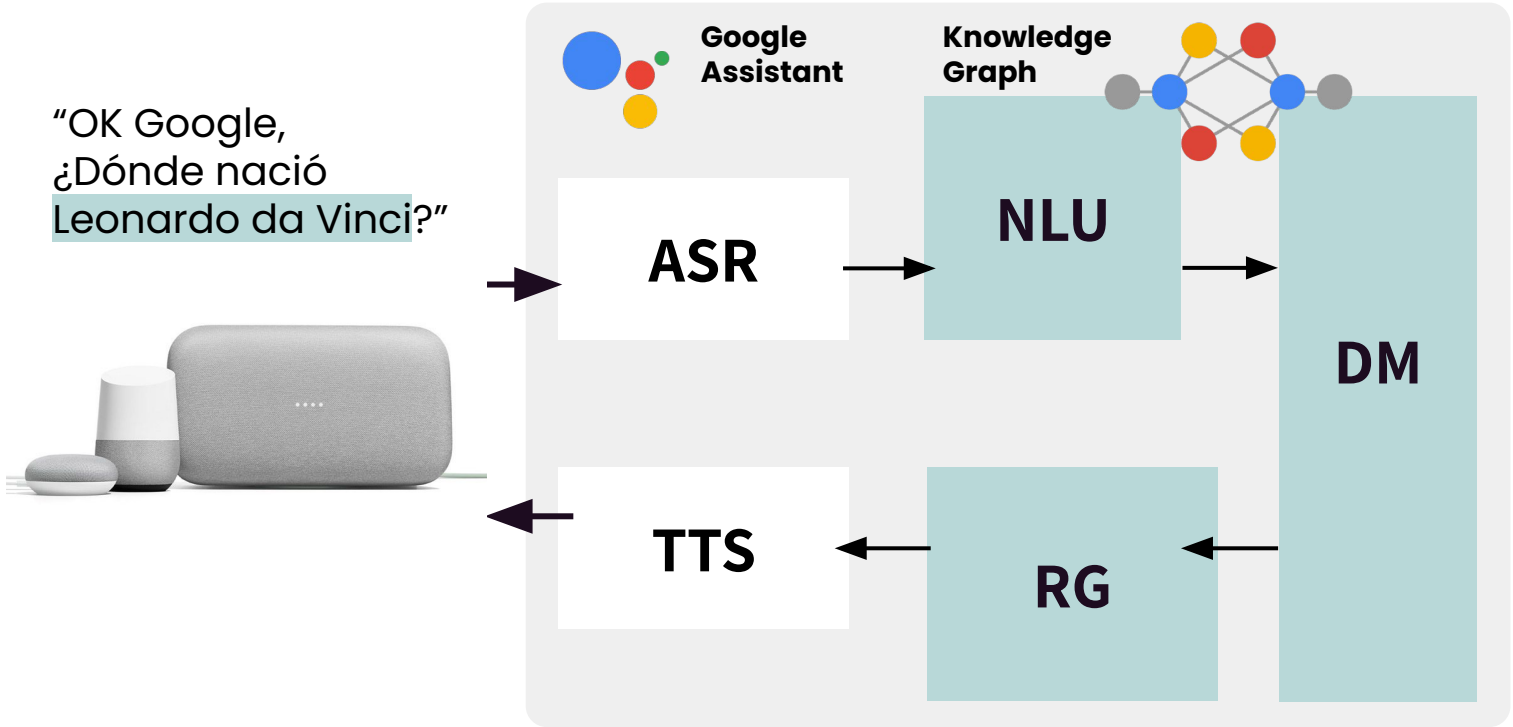
monoceros LABS

**Escribo:**
"abre Juego de Tronos"



# Tecnologías subyacentes (Google Assistant)



# … y mañana?

# ¿Tecnología en fase temprana?

monoceros LABS

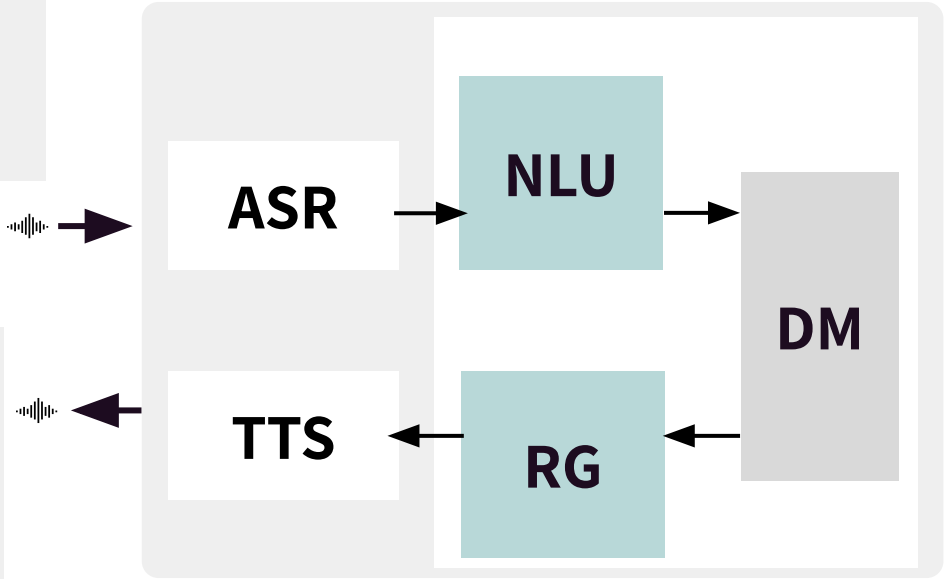
**Advanced features** (*state-of-the-art*, disponibles en algunos asistentes, o en inglés) que van orientadas a buscar la naturalidad en la conversación.

# ¿Tecnología en fase temprana?

**Advanced features** (*state-of-the-art*, disponibles en algunos asistentes, o en inglés) que van orientadas a buscar la naturalidad en la conversación.

**Múltiples intents**: "quiero hacer esto y aquello también"

**Efectos de la voz**: entender y emitir susurros, expresar emociones, personalidad

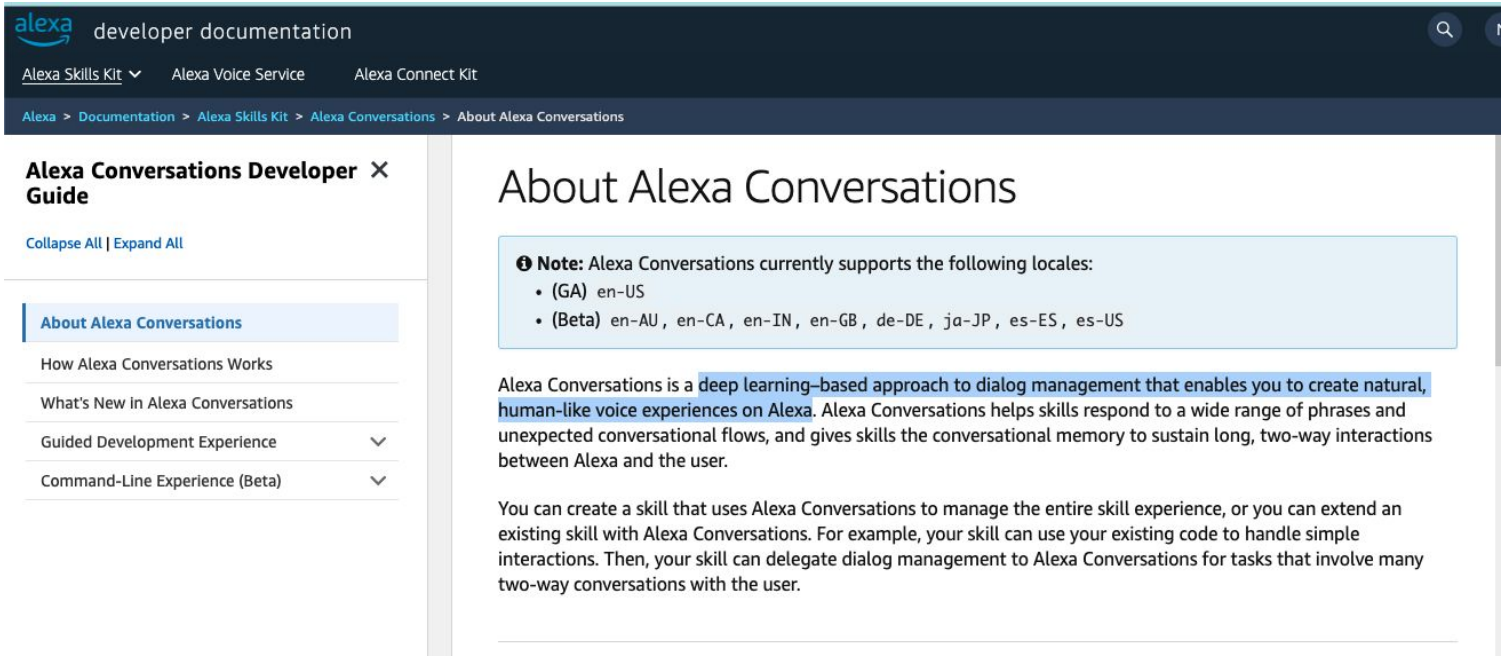



**Contexto**: entender preguntas encadenadas "¿dónde nació?"

**Experiencia personalizada**: usar datos del usuario para recomendar

# ¿Tecnología en fase temprana?

**Advanced features** (*state-of-the-art*, disponibles en algunos asistentes, o en inglés) que van orientadas a buscar la naturalidad en la conversación.



# ¿Tecnología en fase temprana?

monoceros LABS

**Advanced features** (*state-of-the-art*, disponibles en algunos asistentes, o en inglés) que van orientadas a buscar la naturalidad en la conversación.

Use New Alexa Emotions and Speaking Styles to Create a More Natural and Intuitive Voice Experience

Catherine Gao Nov 26, 2019 Share:

Game Skills | Content Skills | Design | News

We're excited to introduce two new Alexa capabilities that will help create a more natural and intuitive voice experience for your customers. Starting today, you can enable Alexa to respond with either a happy/excited or a disappointed/empathetic tone in the US. Emotional responses are particularly relevant to skills in the gaming and sports categories. Additionally, you can have Alexa respond in a speaking style that is more suited for a specific type of content, starting with news and music. Speaking styles are curated text-to-speech voices designed to create a more delightful customer experience for

# ¿Tecnología en fase temprana?

**Advanced features** (*state-of-the-art*, disponibles en algunos asistentes, o en inglés) que van orientadas a buscar la naturalidad en la conversación.

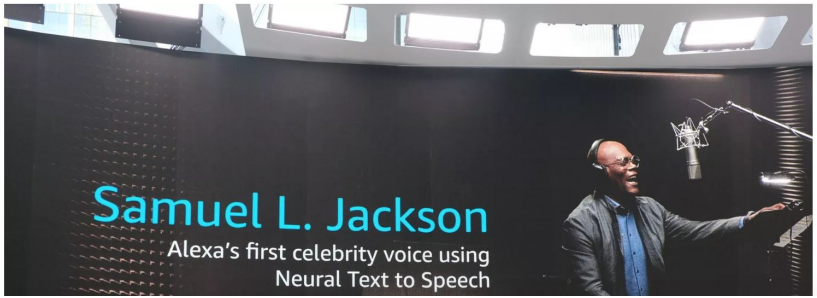
TECH \ AMAZON \ AMAZON ALEXA

## All the new features coming to Alexa, including a new voice, frustration mode, and Samuel L. Jackson

*Also new: bilingual support, and more troubleshooting features like 'frustration detection'*

By Chaim Gartenberg | @cgartenberg | Sep 25, 2019, 1:26pm EDT

SHARE



# ¿Tecnología en fase temprana?

**Advanced features** (*state-of-the-art*, disponibles en algunos asistentes, o en inglés) que van orientadas a buscar la naturalidad en la conversación.

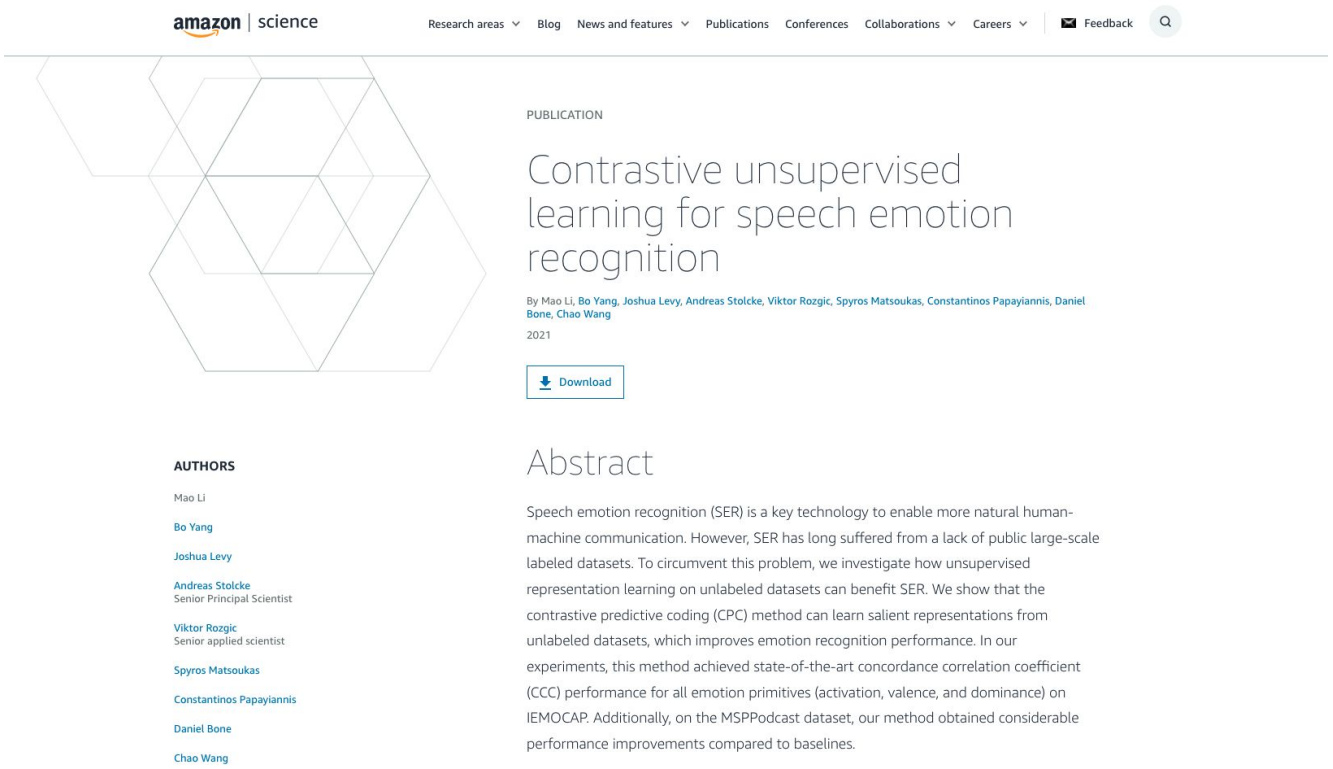
Amazon's neural TTS can model speaking styles with only a few hours of recordings

KYLE WIGGERS @KYLE_L_WIGGERS NOVEMBER 19, 2018 6:55 AM

Above: Amazon Echo Sub

# Y lo que vendrá

**Advanced features** (*state-of-the-art*, aún no disponibles en asistentes).

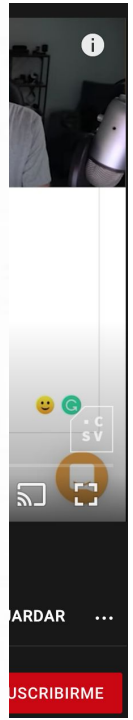


# Y lo que vendrá (... y sus retos)

monoceros LABS

to overc
from da
most co
explicit

The fiel
time. Bi
experie
Amazor
convers
remains



*Gracias por vuestra atención*
*¿Preguntas?*

**Nieves Ábalos (@nieves_as)**

Founder & Chief Product Officer, Monoceros Labs