IA Conversacional ❤️ Python

Nieves Ábalos > @nieves_as

Cofundadora y CPO en Monoceros Labs.

monoceros LABS

Nieves Ábalos Serrano

Chief Product Officer & Cofundadora **Monoceros Labs**

Twitter: **@nieves_as**

LinkedIn: /in/**nievesabalosserrano**

- **Universidad de Granada**: Ingeniera Informática, 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).
- **Comunidad**:
  - **Women in Voice Spain (Lead),**
  - **Alexa Champion,**
  - **AWS Community Builder.**

El presente y futuro de la Inteligencia Artificial está marcado por el lenguaje humano.

Y el lenguaje humano es complejo: símbolos, significado, contexto...

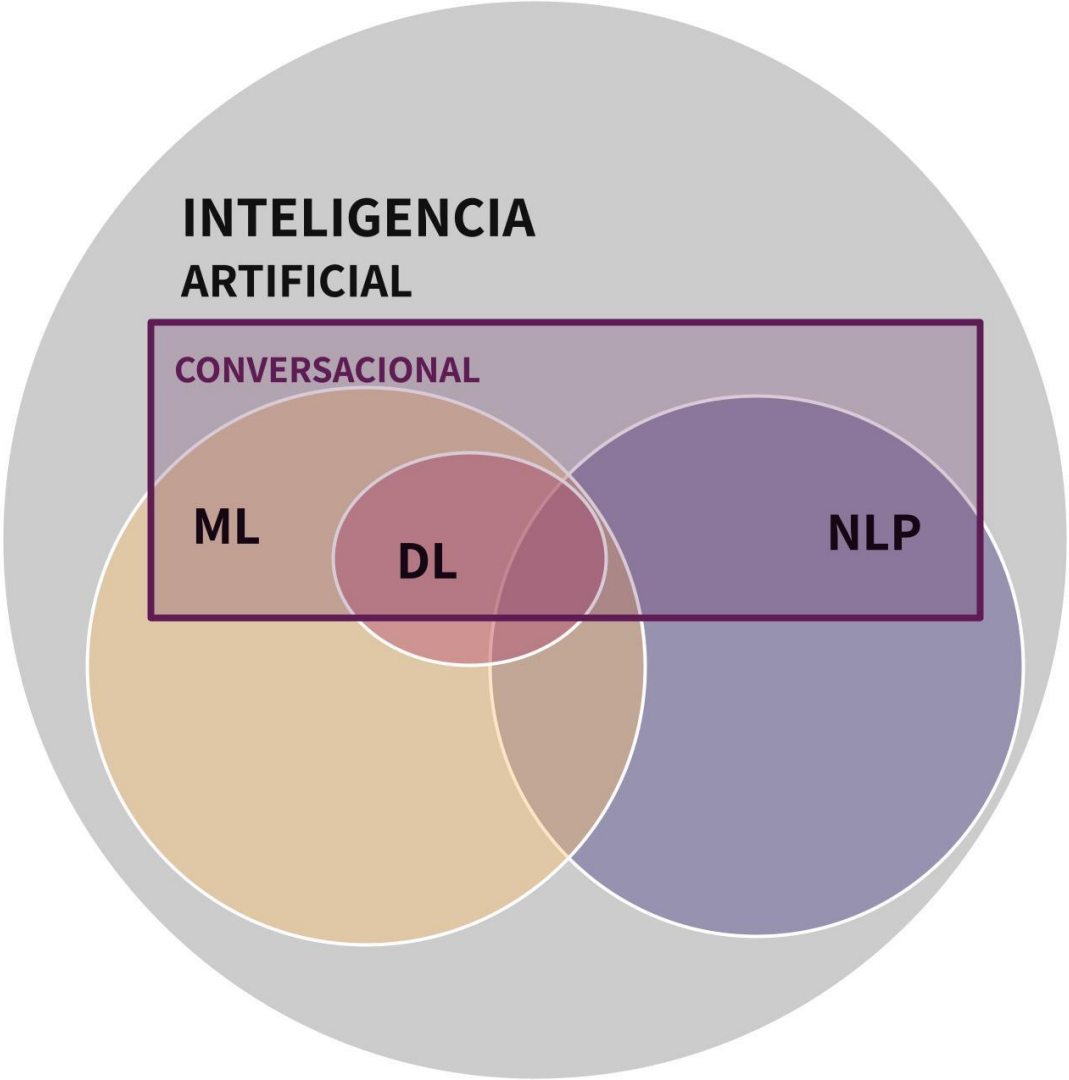
.. el lenguaje humano es impreciso y ambiguo.

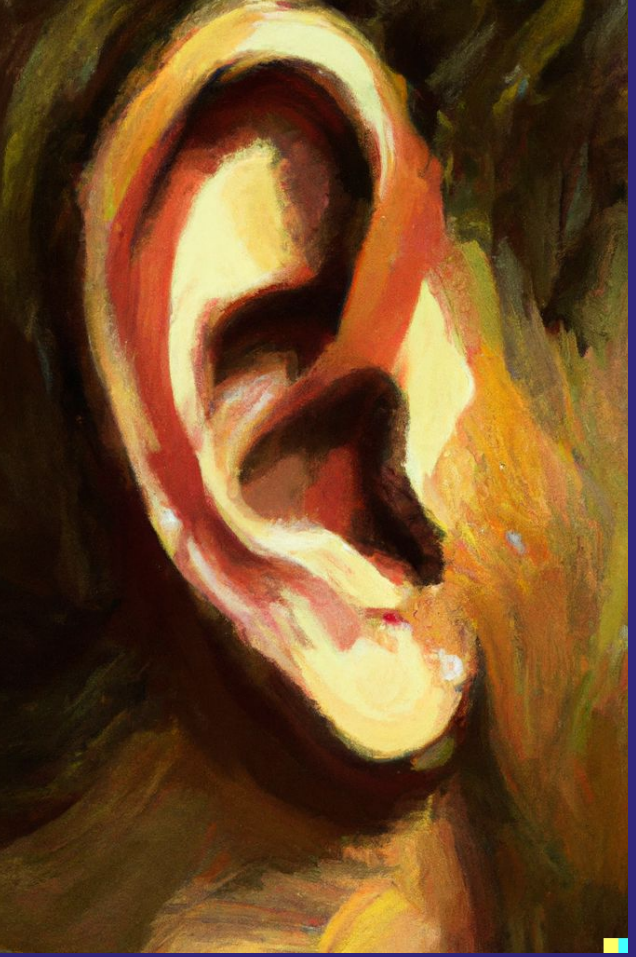
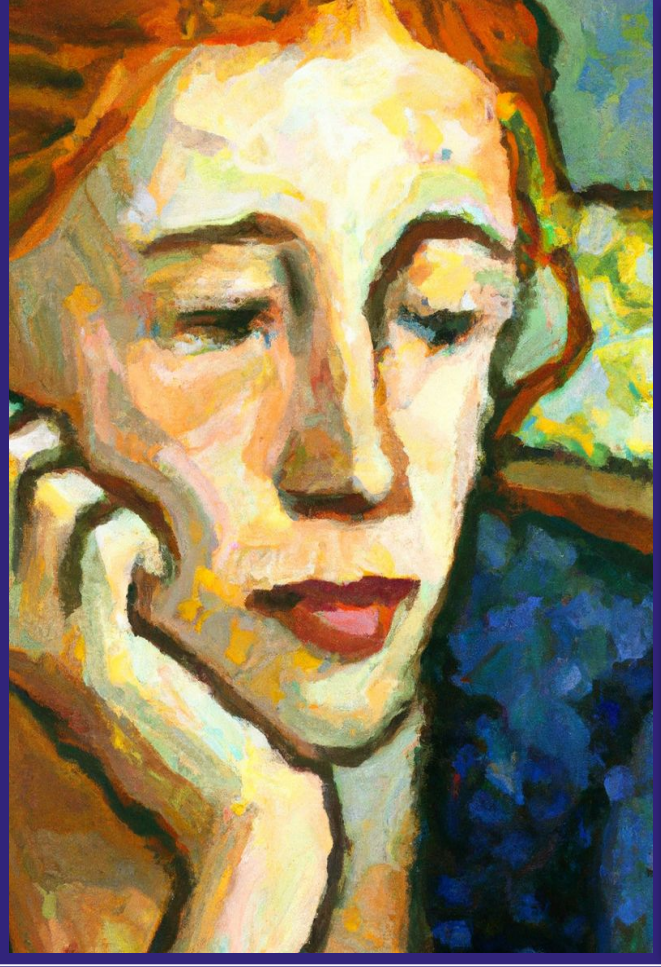
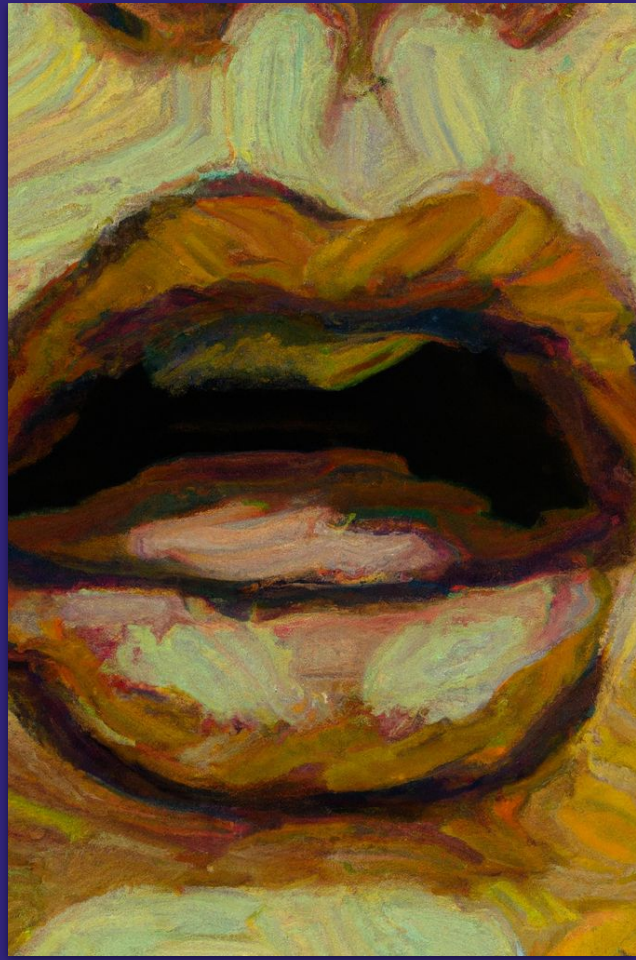
Hablando de ambigüedad

... seguro que habéis escuchado alguno de estos conceptos:

IA
conversacional

🎯 Tecnología
para la
comunicación
con las
personas



Redes Neuronales

... están cambiando las reglas del juego con respecto a lo que podemos construir.

Cómo conseguimos conversaciones con las máquinas más naturales(*) y efectivas

... utilizando Python.

Junio 2022


SAN FRANCISCO
CALIFORNIA
11:26 LIVE
BLAKE LEMOINE | GOOGLE AI ENGINEER
GOOGLE ADVANCEMENTS
ENGINEER CLAIMS COMPANY'S AI MACHINES HAVE THOUGHTS
TALK TV
19:26
THE NEWS DESK

.. ha estado hablando con una máquina:

y ésta es consciente de sí misma, tiene sentimientos, pensamientos.

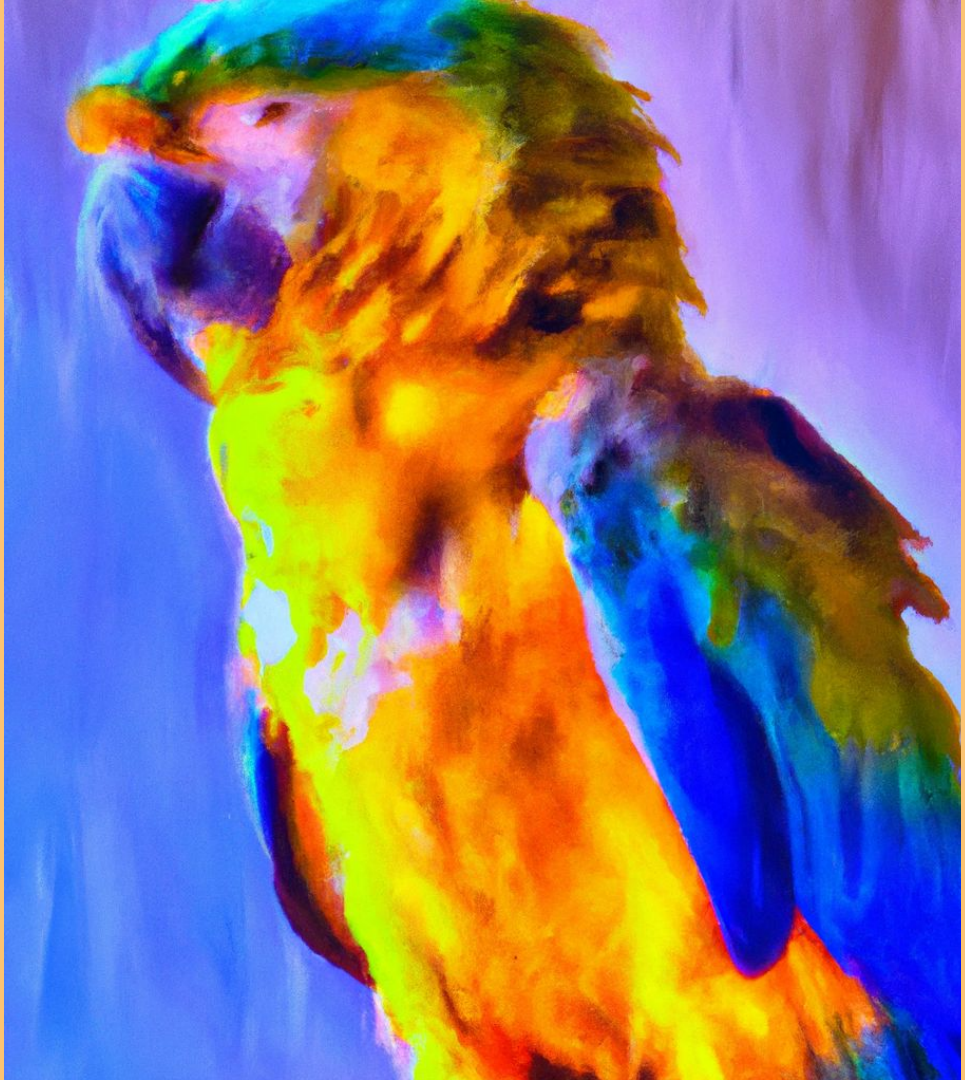
LaMDA (Mayo 2021)

Language Model for Dialogue Applications

blog.google/technology/**ai/lamda**

Es un modelo al que le damos **un objetivo, y buscará conseguirlo a través del diálogo.**

¿LaMDA entiende(*) lo que dice?



LaMDA = modelo del lenguaje:

- **genera** respuestas que nos parece que tienen sentido,
- **aprende posibles respuestas** en base a una **probabilidad**,
- y además esa respuesta la **condicionamos en base al contexto** (lo que le hemos dicho antes).

¿.. un poco loro?

Lemoine habló con un modelo del lenguaje cuyo objetivo era convencerlo de que tenía consciencia...

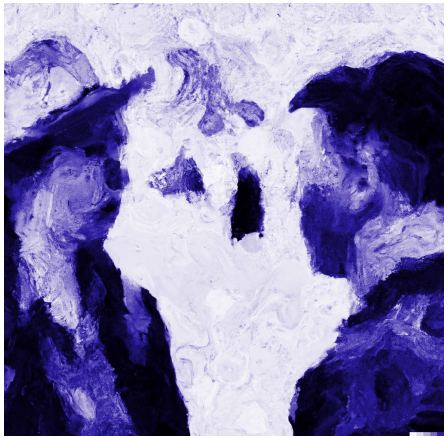
… y la consciencia es una cualidad humana.

Antropo-morfismo

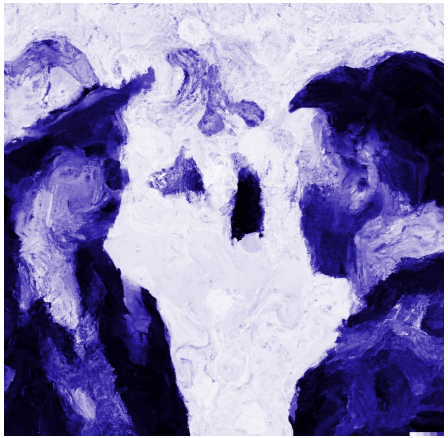
Dar **forma** o **cualidades humanas** a lo que no lo es.

¿Qué cualidades humanas buscamos en la tecnología conversacional?

COMUNICAR EN LENGUAJE NATURAL

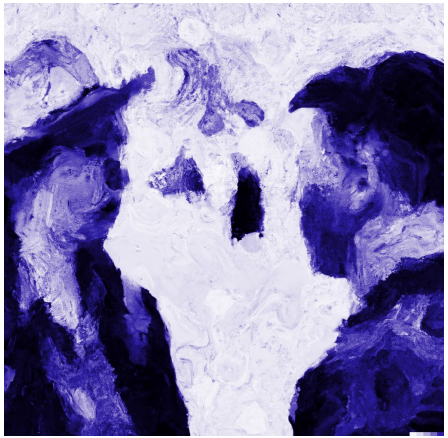


COMUNICAR EN LENGUAJE NATURAL



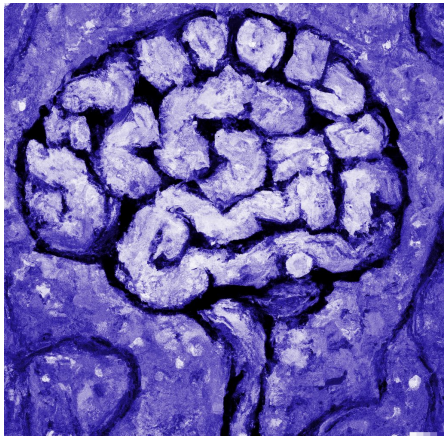
CREAR UN VÍNCULO

COMUNICAR EN LENGUAJE NATURAL

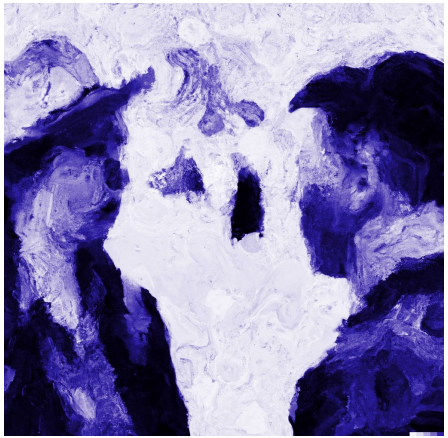


CREAR UN VÍNCULO

INTELIGENCIA

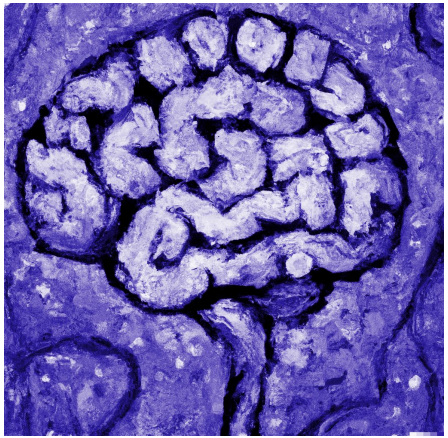


COMUNICAR EN LENGUAJE NATURAL



CREAR UN VÍNCULO

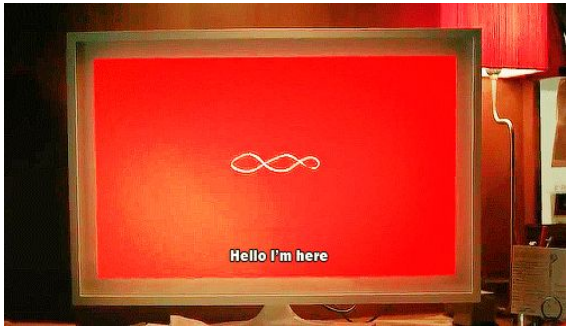
INTELIGENCIA

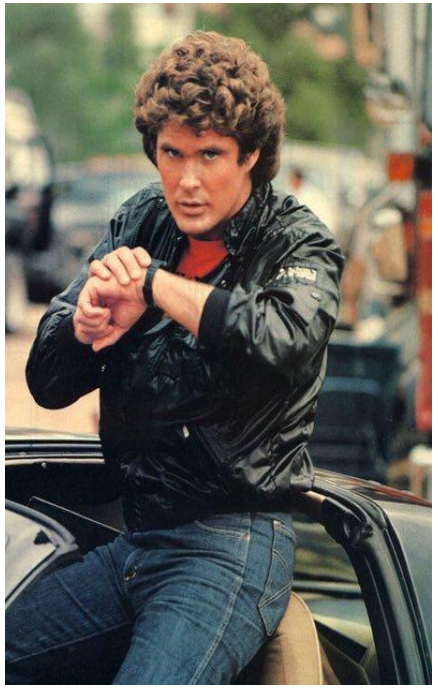


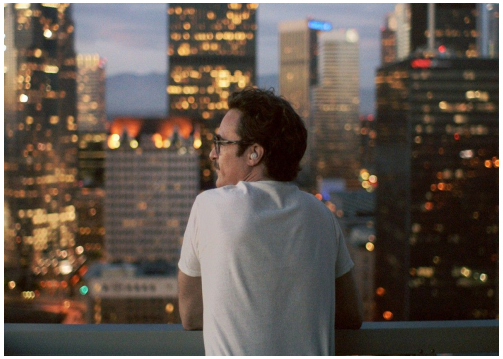
CONFIANZA

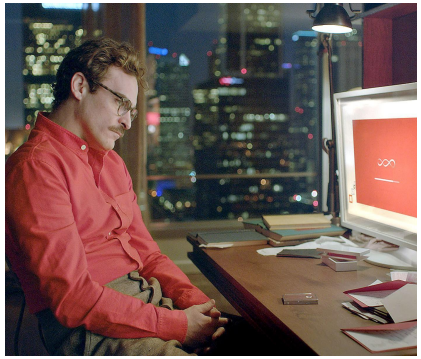
¿Expectativas altas...?

¡Ay, la ciencia ficción!







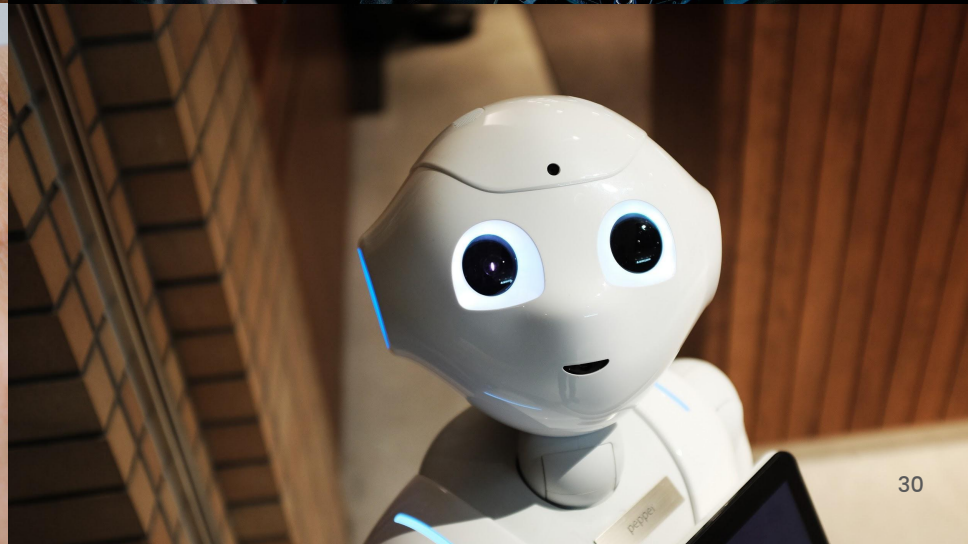


Y luego, la realidad...

Vamos con la realidad: qué podemos construir con la tecnología que tenemos disponible (y qué limitaciones tiene)


1:22
Weird Thing to Know
1:22


ACTIVITY
Weekly
Summary
11.03
Happy
New Year,
Fikri!



.. pero no todas las conversaciones son igual de fáciles de construir.

La complejidad según el tipo de conversación...

Closed-domain (un dominio concreto)

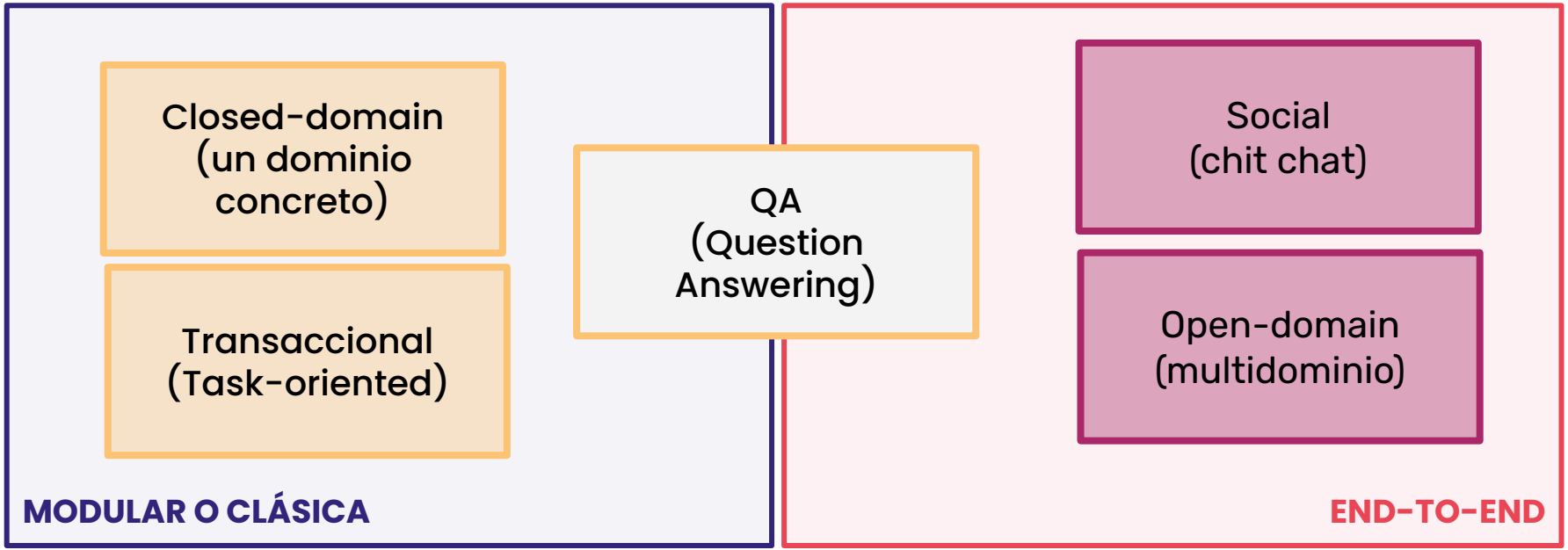
Transaccional (Task-oriented)

QA (Question Answering)

Social (chit chat)

Open-domain (multidominio)

.. y (generalizando) según las piezas de la arquitectura.



Arquitectura modular o clásica.

Los módulos o fases: interfaz de voz

Alexa (Wake word), pon Rosalía en Spotify

Aquí tienes Rosalía...

Los módulos o fases: interfaz de voz

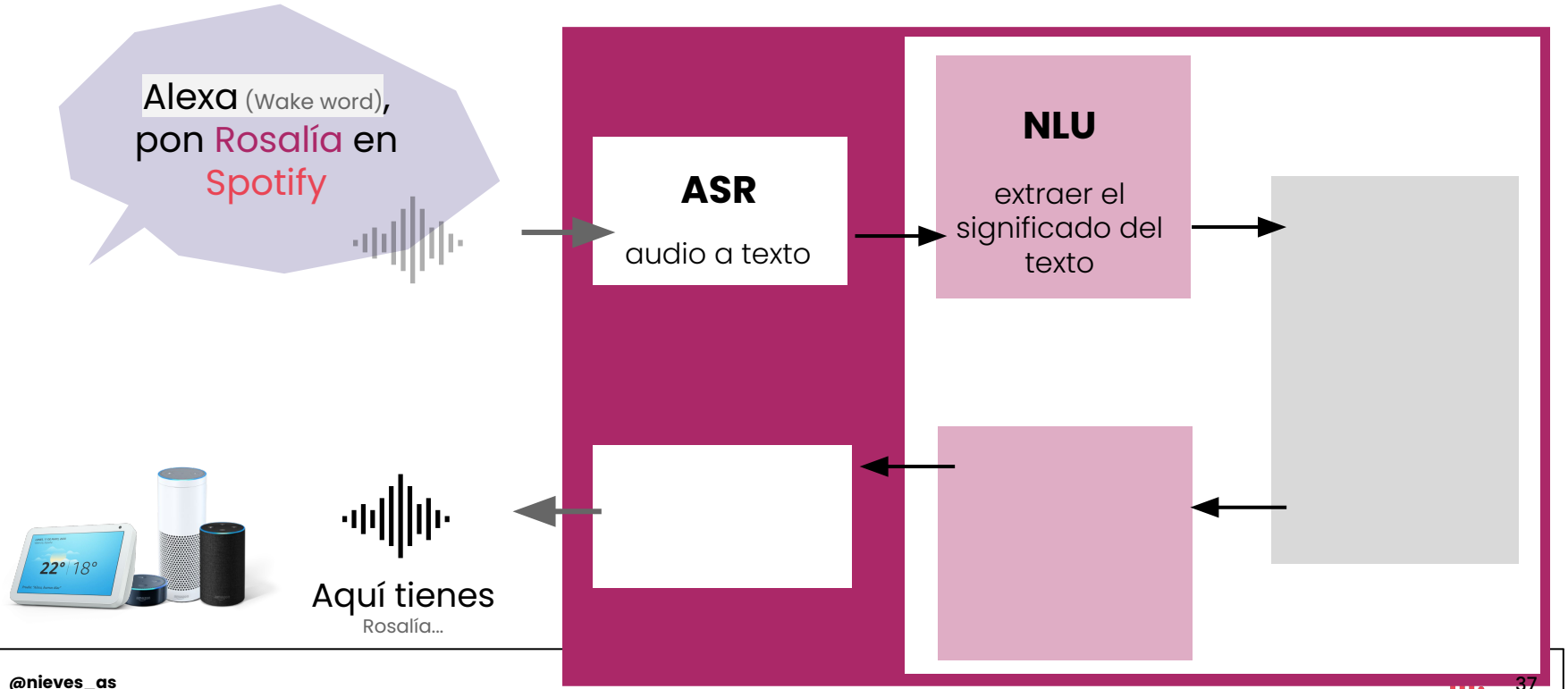
Alexa (Wake word), pon Rosalía en Spotify

Aquí tienes
Rosalía...

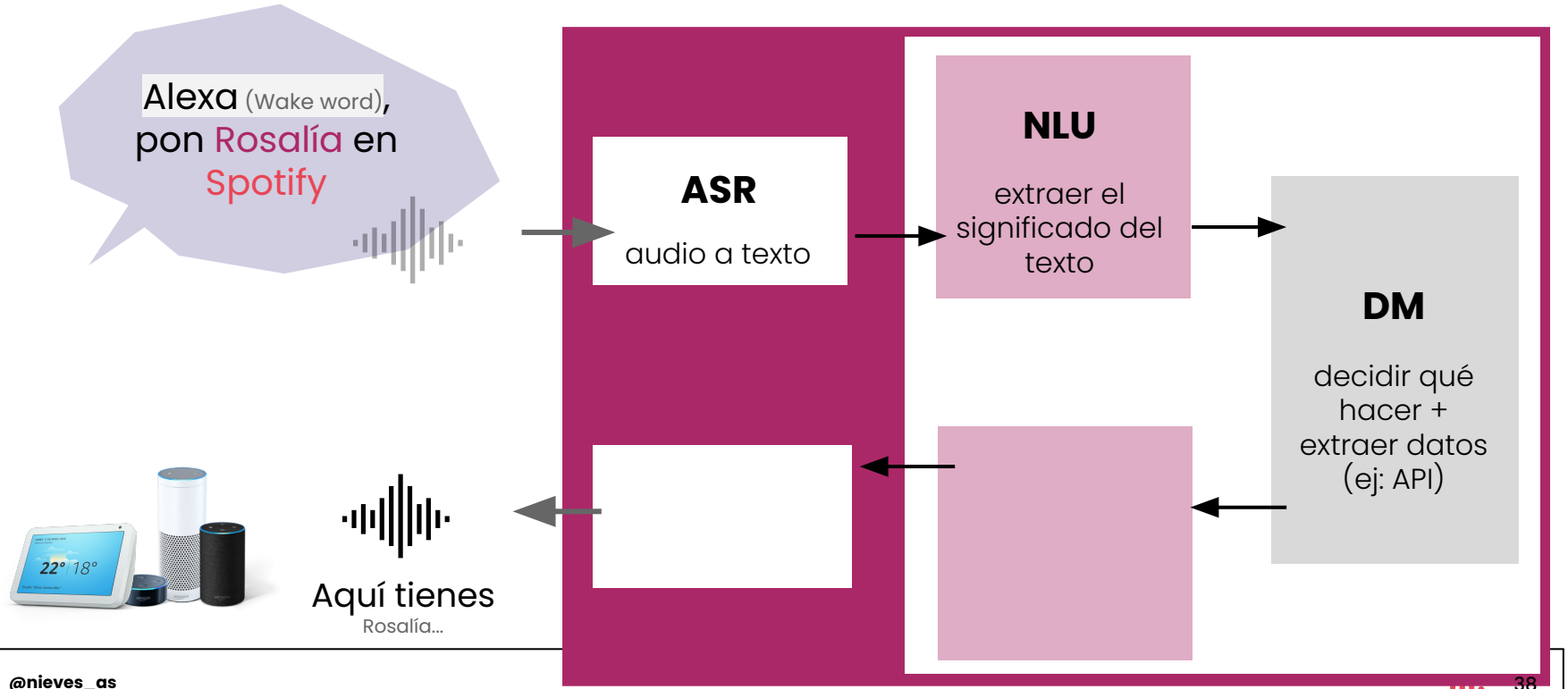
ASR

audio a texto

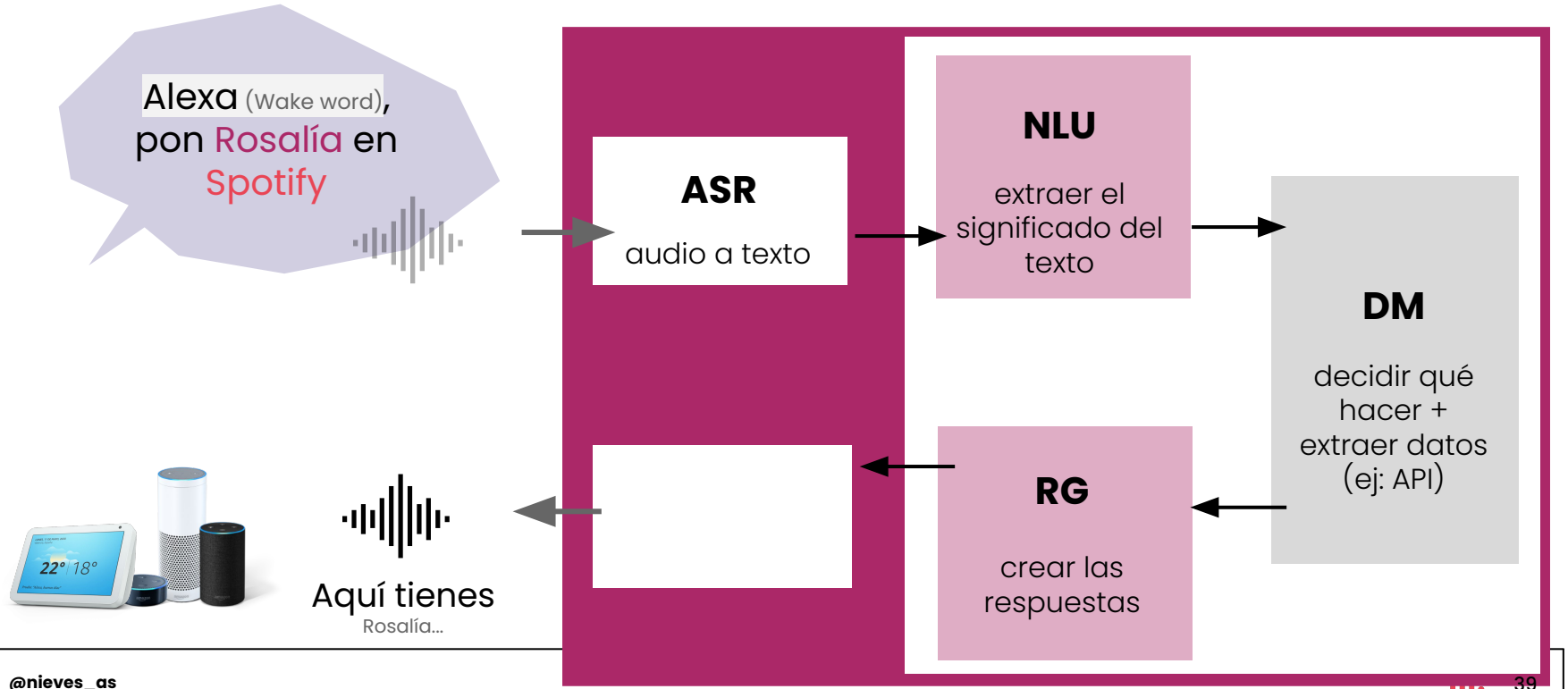
Los módulos o fases: interfaz de voz



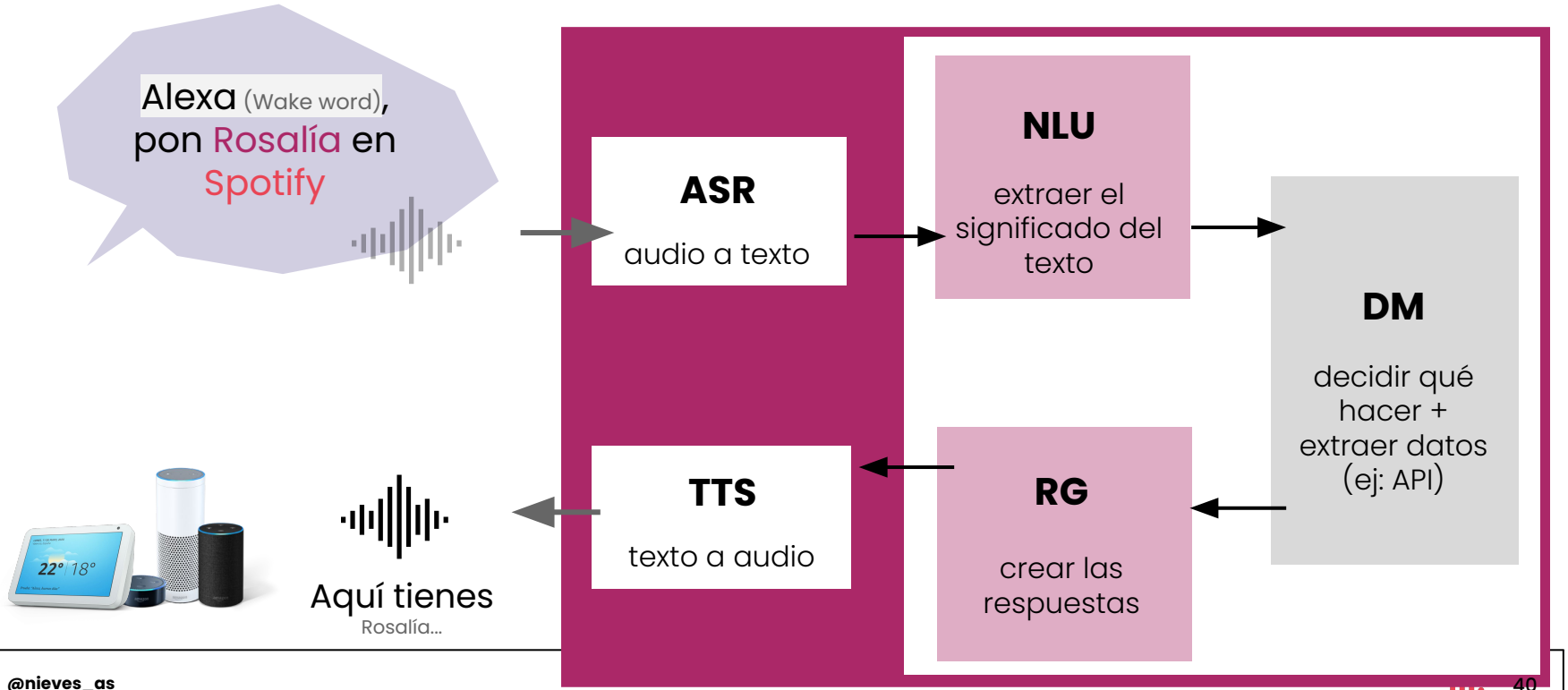
Los módulos o fases: interfaz de voz



Los módulos o fases: interfaz de voz

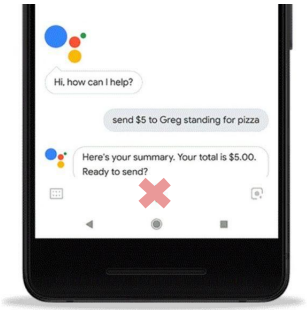


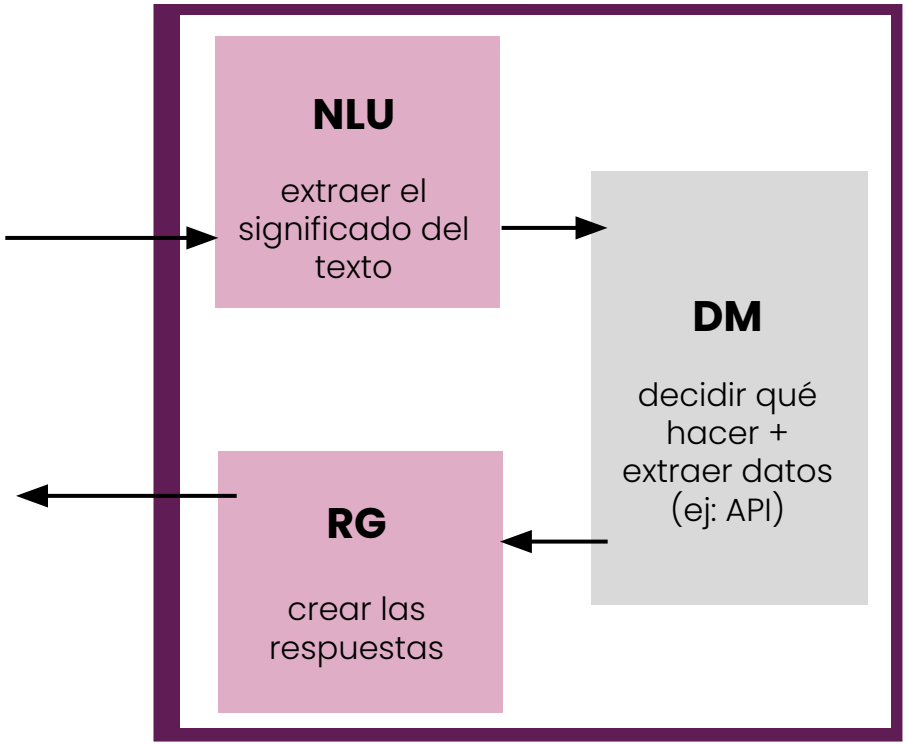
Los módulos o fases: interfaz de voz



Los módulos o fases: interfaz de texto

Escribo:
"abre Juego de Tronos"

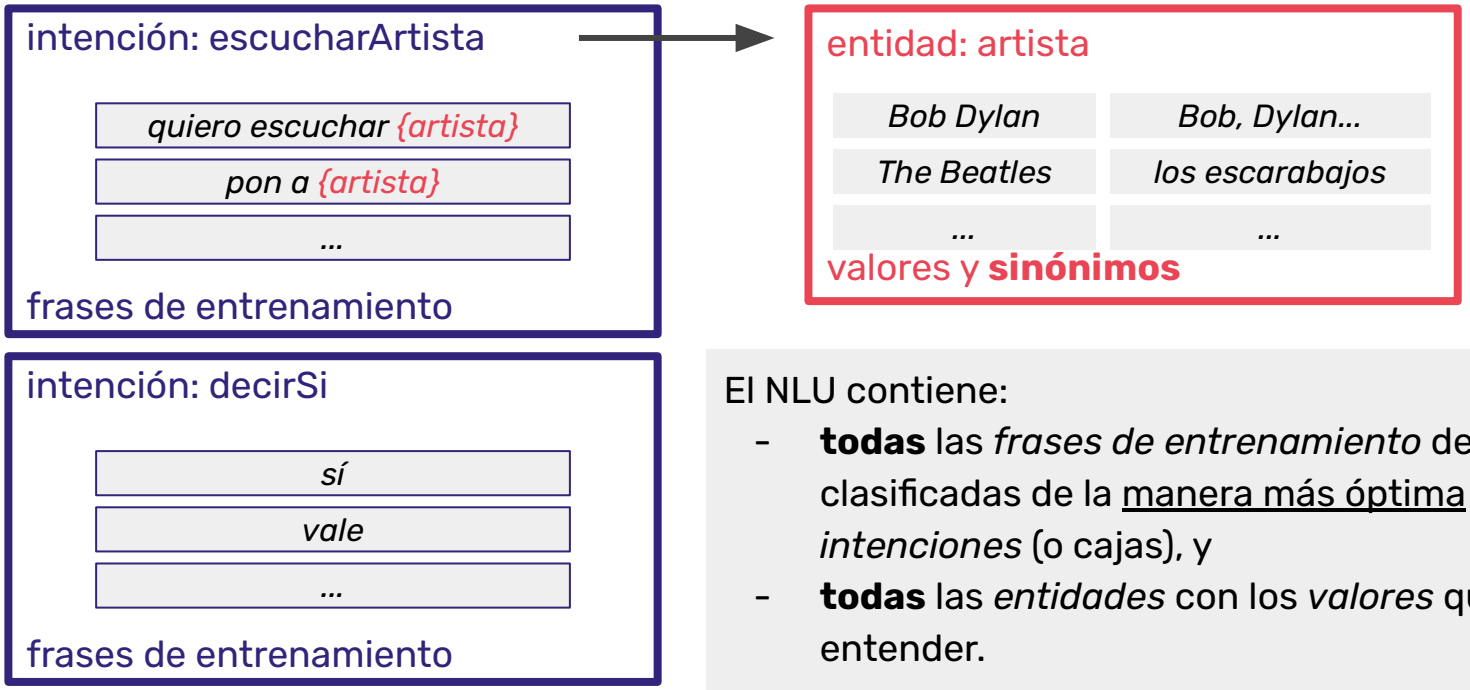




NLU

Lo más importante: intentar entender el significado.
Esta pieza está **presente** en casi todas **las plataformas de voz y texto.**

El NLU en estas plataformas es un modelo que contiene...



El NLU contiene:

- **todas** las *frases de entrenamiento* del dominio, clasificadas de la manera más óptima en las *intenciones* (o cajas), y
- **todas** las *entidades* con los *valores* que podemos entender.

Otras herramientas
NLU / NLP

spaCy

NLP AVANZADO con spaCy

Capítulo 1: Encontrando palabras, frases, nombres y conceptos

Este capítulo te mostrará las bases del procesamiento de texto con spaCy. Aprenderás acerca de las estructuras de datos, cómo trabajar con pipelines entrenados y cómo usar eastas estructuras y pipelines para predecir características lingüísticas en tu texto.

Capítulo 2: Análisis de datos a gran escala con spaCy

En este capítulo usarás tus nuevas habilidades para extraer información específica de grandes volúmenes de texto. Aprenderás a sacarle el mayor provecho a las estructuras de datos de spaCy y cómo combinar los enfoques estadísticos y basados en reglas de manera efectiva para el análisis de texto.

Capítulo 3: Pipelines de procesamiento

En este capítulo aprenderás todo lo que necesitas saber sobre el pipeline de procesamiento de spaCy. Aprenderás lo que sucede cuando procesas un texto, cómo escribir tus propios componentes y añadirlos al pipeline y cómo usar atributos personalizados para añadir tus propios metadatos a los documentos, spans y tokens.

Capítulo 4: Entrenando un modelo de red neuronal

wit.ai

Duckling
Github Repository
Try it out
Introduction
Limitations
Getting Started
Implementation
Extending Duckling
Debugging
Contributing
FAQ
by wit.ai

Duckling

The linguist (>) that parses text into structured data

Try it out with a date/time

e.g. tomorrow at 6am EN Try me!

Introduction

Duckling is a Clojure library that parses text into structured data:

```
"the first Tuesday of October" => {:value "2014-10-07T00:00:00.000-07:00"
                                   :grain :day}
```

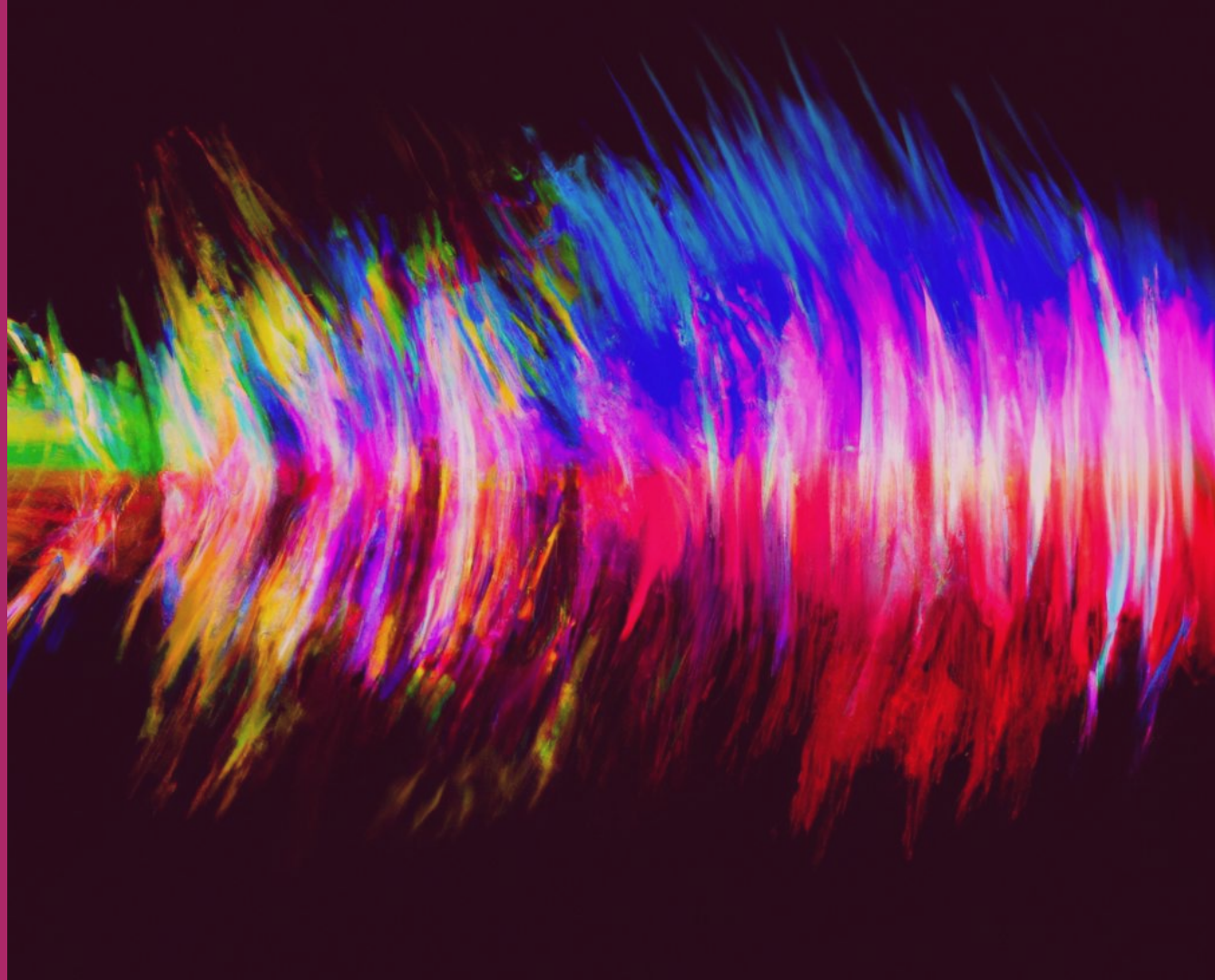
See our blog post announcement for more context.

Duckling is shipped with modules that parse temporal expressions in English, Spanish, French, Italian and Chinese (experimental, thanks to Zhe Wang). It recognizes dates and times described in many ways:

- *today at 5pm*
- *2014-10-01*
- *the last Tuesday of October 2012*
- *twenty five minutes ago*
- *the day before labor day 2020*
- *June 10-11* (interval)
- third monday after christmas 1980

TTS

De *Loquendo* a voces neurales.



Voces sintéticas

TEXTO

"Buenos días, esta es nuestra última clase sobre interfaces de voz"

MODELO DE VOZ SINTÉTICA

KITTS: clonación y voces que no existen.

Voz clonada
Nieves

Voz que no existe
www.**estavoznoexiste**.com

La primera voz sin género

ASR

Para transcribir de voz a texto.
También llamado *speech to text*.

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).

Whisper
(OpenAI, 2022)

open-source

Introducing Whisper

We've trained and are open-sourcing a neural net called Whisper that approaches human level robustness and accuracy on English speech recognition.

September 21, 2022
7 minute read

READ PAPER

VIEW CODE

VIEW MODEL CARD

Whisper examples: Speed talking

REVEAL TRANSCRIPT

Whisper is an automatic speech recognition (ASR) system trained on 680,000 hours of multilingual and multitask supervised data collected from the web. We show that the use of such a large and diverse dataset leads to improved

Whisper (OpenAI, 2022)

```
audio = "/home/carlos/Música/7db883e9-ab55-47df-9924-c1cb63a8941d.mp3"

result = transcribe_audio(audio)

ipd.display(ipd.Audio(audio, rate=22050, normalize=True))
print(result["text"])
print()
write_vtt(result["segments"])
```

```
Detected language: spanish
```


```
 Es hora de crear conversaciones. Somos Monoceros, un estudio de innovación especializado en estrategia, diseño conversacional y tec
nologías del habla. Creamos aplicaciones de voz para asistentes como Amazon Alexa y Google Assistant. Además, tenemos un proyecto de
investigación tecnológica para crear voces sintéticas más naturales y expresivas en español. El producto se llamará Voces.ai y lo pu
blicaremos este año 2022. Esta voz que escuchas está generada con esta tecnología. Para conseguir este resultado solo hemos necesita
do 200 frases de nieve.

Subtítulos WEBVTT

00:00.000 --> 00:02.000
 Es hora de crear conversaciones.

00:02.000 --> 00:10.000
 Somos Monoceros, un estudio de innovación especializado en estrategia, diseño conversacional y tecnologías del habla.

00:10.000 --> 00:15.000
 Creamos aplicaciones de voz para asistentes como Amazon Alexa y Google Assistant.

00:15.000 --> 00:23.000
 Además, tenemos un proyecto de investigación tecnológica para crear voces sintéticas más naturales y expresivas en español.

00:23.000 --> 00:29.000
 El producto se llamará Voces.ai y lo publicaremos este año 2022.

00:29.000 --> 00:33.000
 Esta voz que escuchas está generada con esta tecnología.

00:33.000 --> 01:01.000
 Para conseguir este resultado solo hemos necesitado 200 frases de nieve.
```

Asistentes y plataformas

Propietarias y open-source

Plataformas para crear apps o asistentes

Propietarias

amazon alexa

Google Assistant
(*) app actions

Bot Framework SDK

Dialogflow

Voiceflow

IBM **Watson** Assistant

wit.ai

Open-source

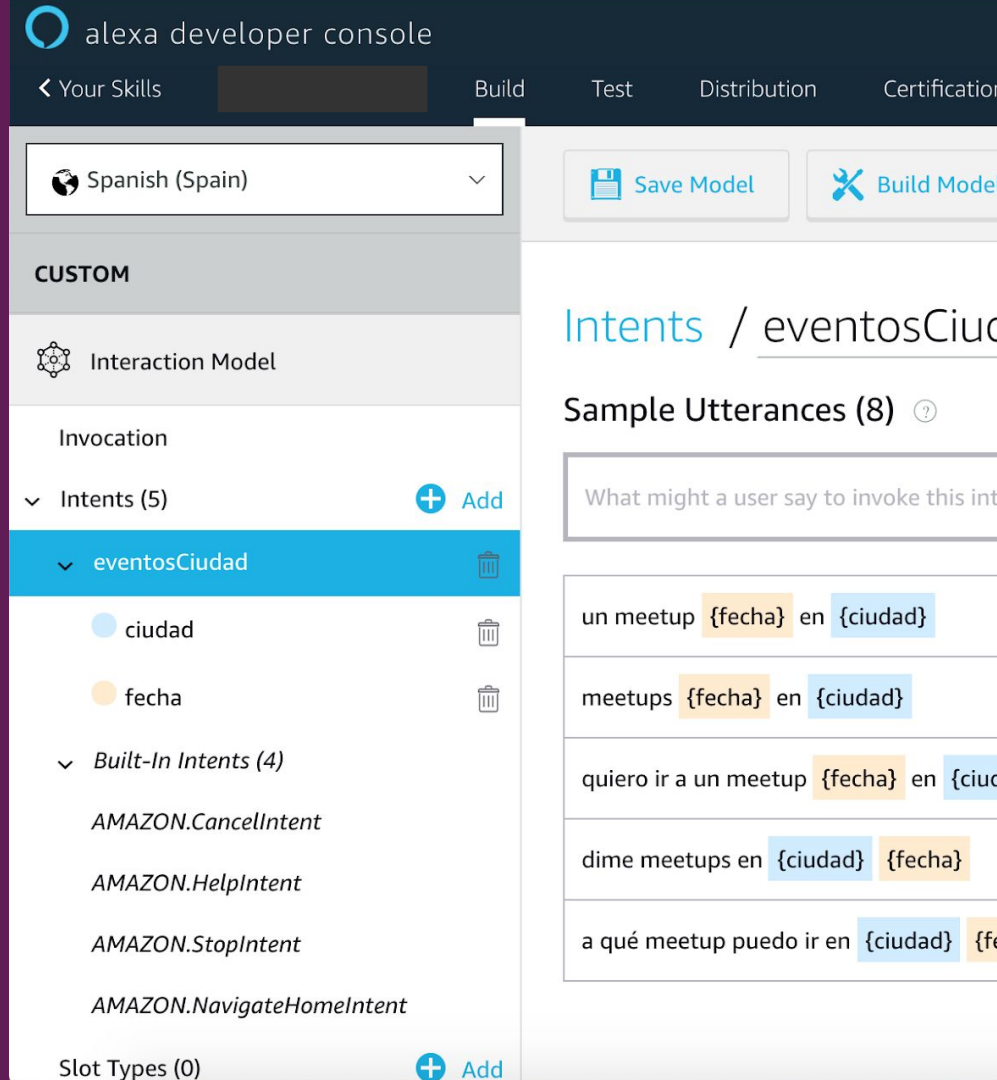
RASA

MYCROFT AI

Home Assistant

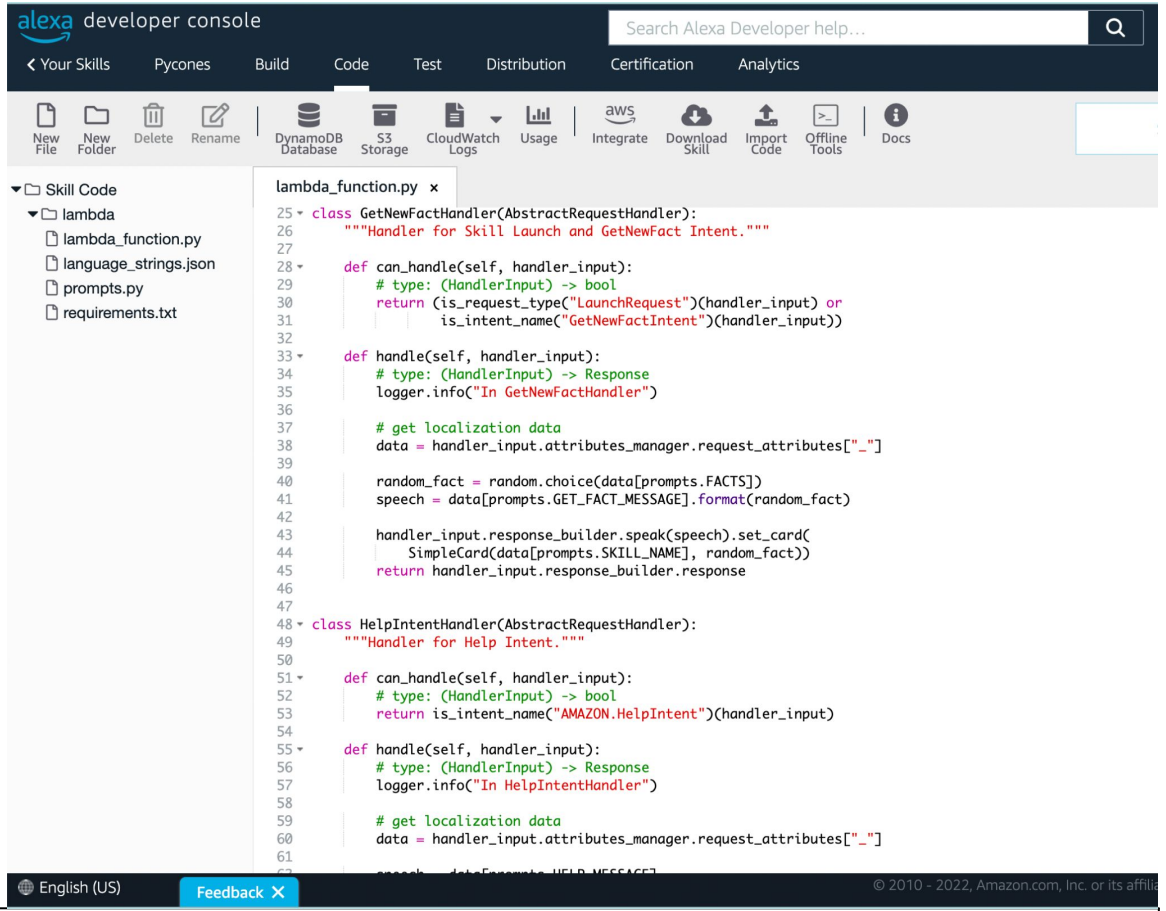
Alexa
(Hosted-Skill)

Crea una Alexa Skill con Python desde la Alexa Developer Console.

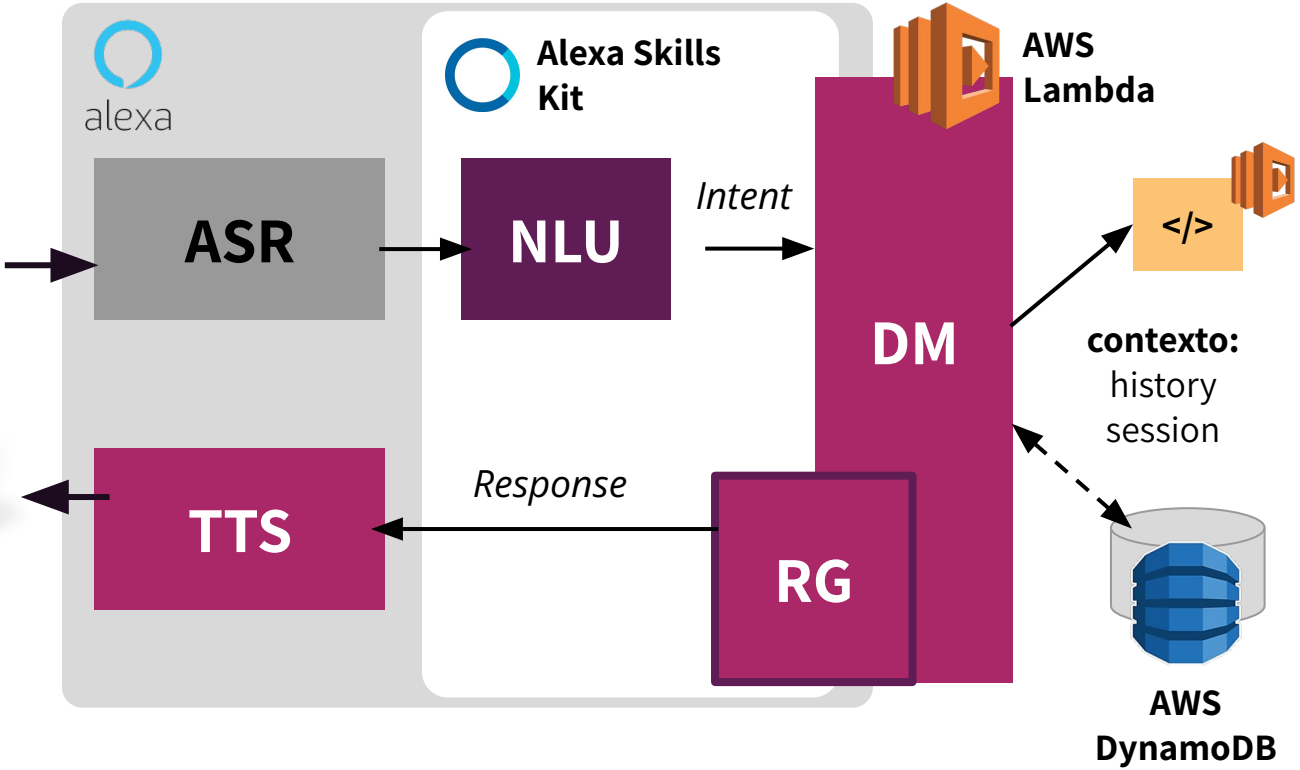


Alexa Hosted-Skill

Templates en Python.



Digo:
“Alexa,
abre **Veo Veo**”



Rasa

Crea un chatbot con tecnología con los mejores modelos y algoritmos del estado del arte, y open source.

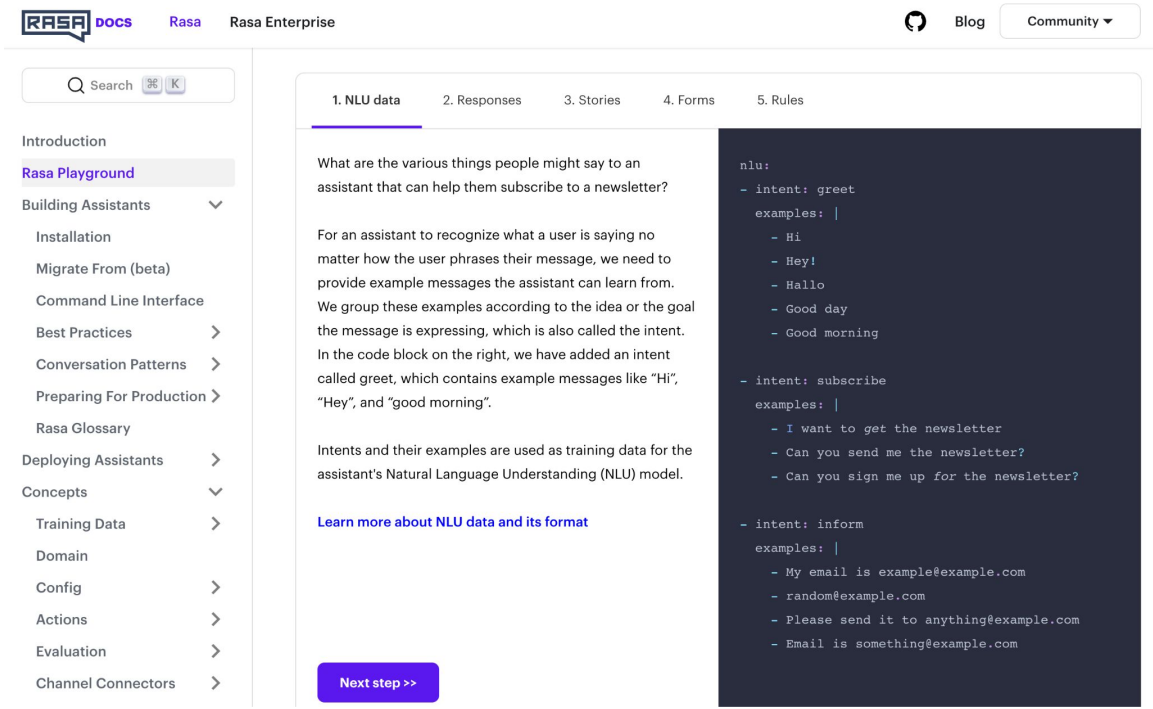
actions.py

```python
# This files contains your custom actions which can
# custom Python code.
#
# See this guide on how to implement these action:
# https://rasa.com/docs/rasa/core/actions/#custom-ac


# This is a simple example for a custom action which

from typing import Any, Text, Dict, List

from rasa_sdk import Action, Tracker
from rasa_sdk.executor import CollectingDispatcher
from rasa_sdk.events import SlotSet


class ActionFacilitySearch(Action):

    def name(self) -> Text:
        return "action_facility_search"

    def run(self, dispatcher: CollectingDispatcher,
            tracker: Tracker,
            domain: Dict[Text, Any]) -> List[Dict[Te

        facility = tracker.get_slot("facility_type")
        address = "300 Hyde St, San Francisco"
        dispatcher.utter_message("Here is the addres

        return [SlotSet("address", address)]
```

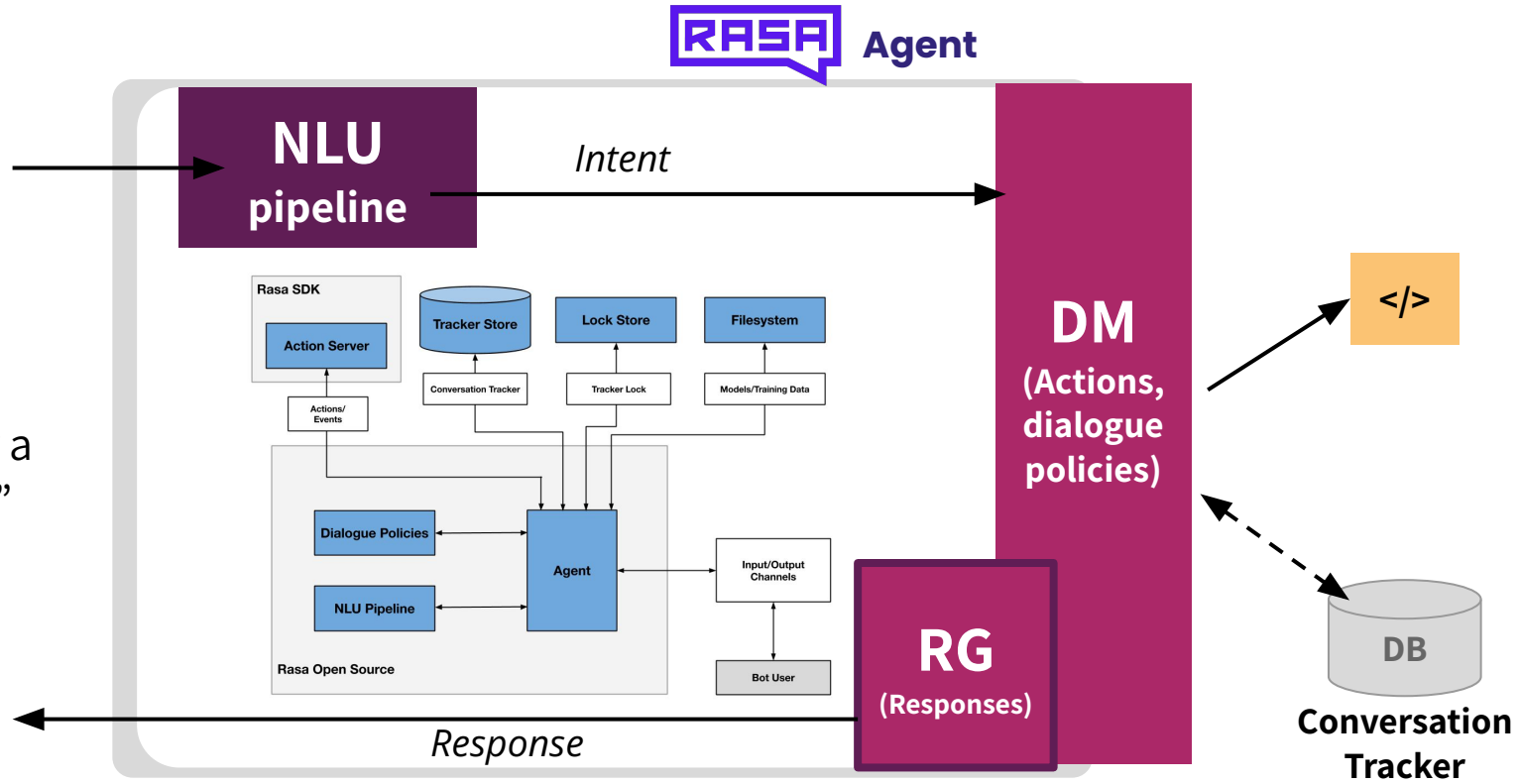
stories.md

```
## hospital search
* greet
  - utter_how_can_
* search_provider{
"location": "San F
  - action_facilit
  - slot{}
* thanks
  - utter_goodbye

## hospital search
* greet
  - utter_how_can_
* search_provider{
  - utter_ask_loca
* inform{"location
  - action_facilit
* thanks
  - utter_goodbye

## happy path
* greet
  - utter_greet
* mood_great
  - utter_happy

## sad path 1
* greet
  - utter_greet
* mood_unhappy
  - utter_cheer_up
  - utter_did_that
* affirm
  - utter_happy
```

RASA Playground (tutorial)



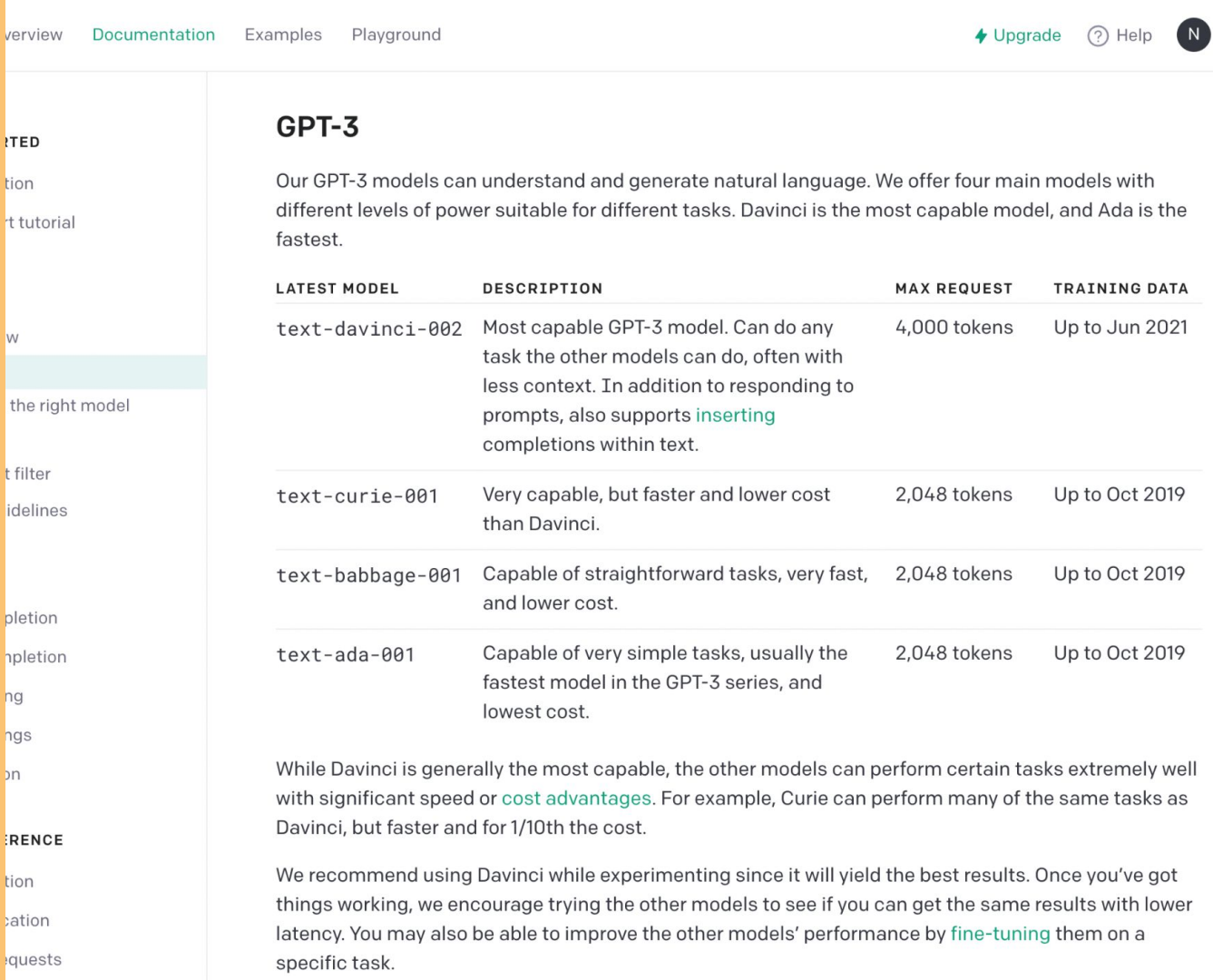
escribo:
“quiero ir a **Granada**”



Arquitectura end-to-end.

GPT-3
(OpenAI, 2020)

Modelos del lenguaje con múltiples habilidades.



GPT-3

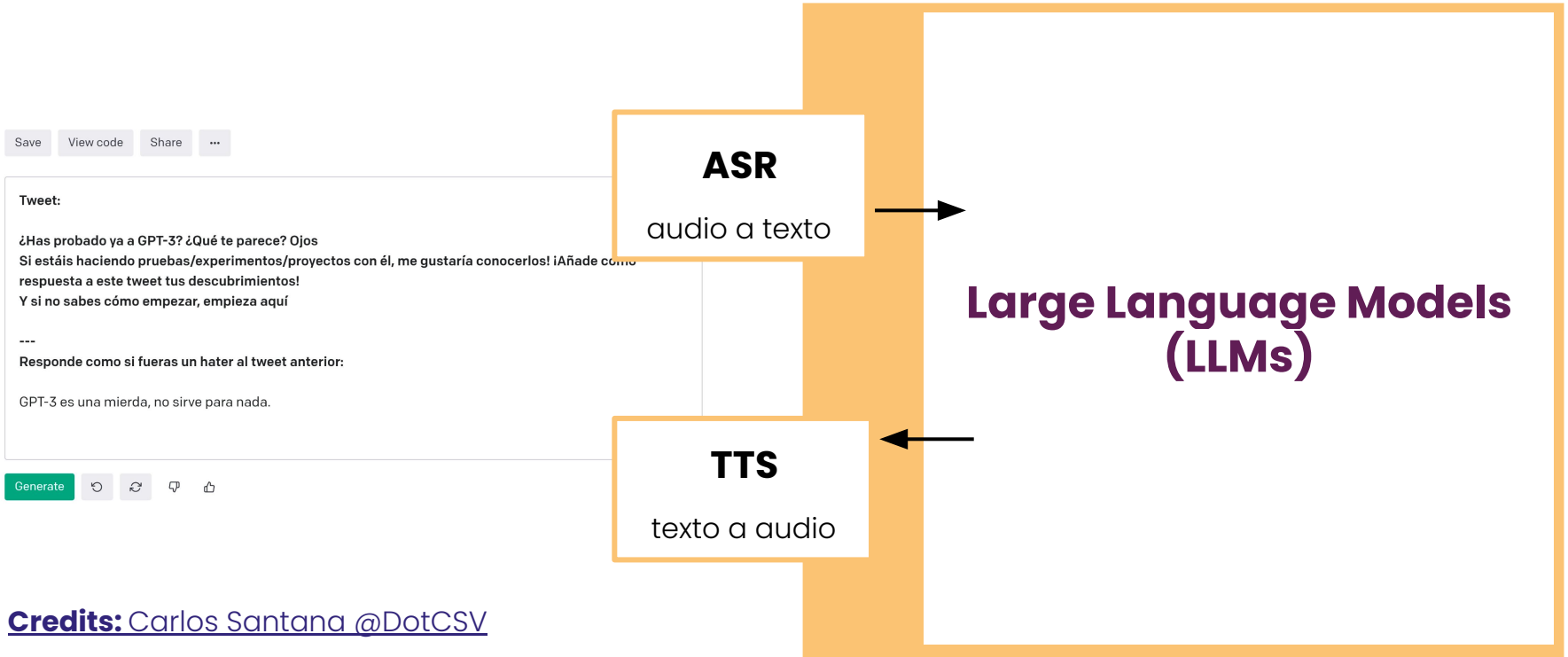
Our GPT-3 models can understand and generate natural language. We offer four main models with different levels of power suitable for different tasks. Davinci is the most capable model, and Ada is the fastest.

LATEST MODEL	DESCRIPTION	MAX REQUEST	TRAINING DATA
`text-davinci-002`	Most capable GPT-3 model. Can do any task the other models can do, often with less context. In addition to responding to prompts, also supports inserting completions within text.	4,000 tokens	Up to Jun 2021
`text-curie-001`	Very capable, but faster and lower cost than Davinci.	2,048 tokens	Up to Oct 2019
`text-babbage-001`	Capable of straightforward tasks, very fast, and lower cost.	2,048 tokens	Up to Oct 2019
`text-ada-001`	Capable of very simple tasks, usually the fastest model in the GPT-3 series, and lowest cost.	2,048 tokens	Up to Oct 2019

While Davinci is generally the most capable, the other models can perform certain tasks extremely well with significant speed or cost advantages. For example, Curie can perform many of the same tasks as Davinci, but faster and for 1/10th the cost.

We recommend using Davinci while experimenting since it will yield the best results. Once you've got things working, we encourage trying the other models to see if you can get the same results with lower latency. You may also be able to improve the other models' performance by fine-tuning them on a specific task.

Open domain conversations: end-to-end



Credits: Carlos Santana @DotCSV

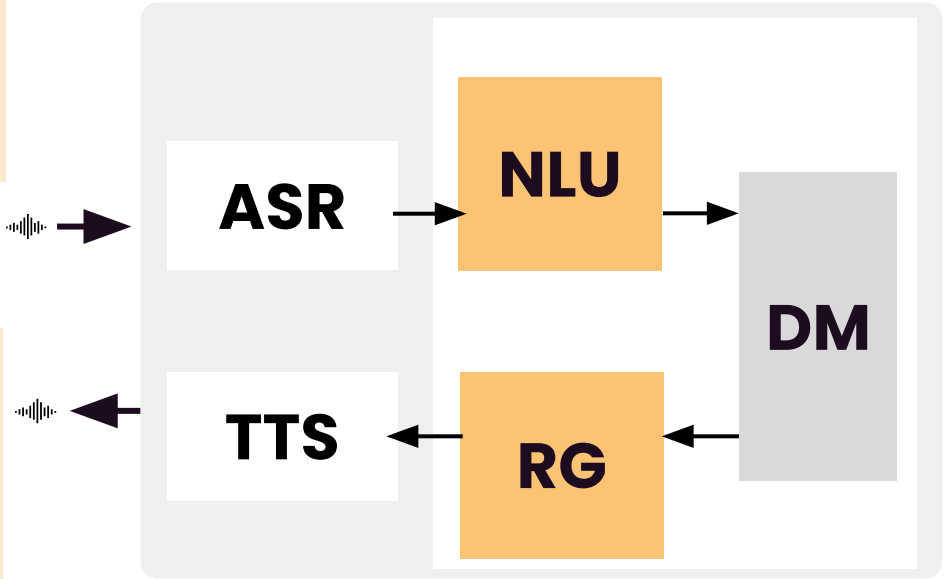
Y qué más vendrá...

Más estado del arte de la IA conversacional.

Más SOTA (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"



Contexto: entender preguntas encadenadas "¿dónde nació?"

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

Experiencia personalizada: usar datos del usuario para recomendar

En general, estos son los retos de la IA conversacional

Entender el CONTEXTO y mantenerlo

APRENDIZAJE

Activo y a largo plazo

Multimodalidad y multitarea

Conversaciones

OPEN-DOMAIN

Estos avances tecnológicos son apasionantes ❤️

... hay
cualidades NO humanas
de la IA conversacional
que nos dan valor.. 💎

… tecnología que nos ayuda, que nos hace la vida más fácil 💎

no nos distraigamos antropomorfizando, dando cualidades humanas a la tecnología..

Los aviones son probados según su habilidad para volar, no comparándolos con aves.

“IA: un enfoque moderno” - Russell y Norvig

¡Gracias por vuestra atención!

NIEVES ÁBALOS

Cofundadora y CPO @ Monoceros Labs

✉️ nieves@monoceroslabs.com

🐦 @nieves_as

💼 /in/nievesabalosserrano