

# TALLER PROCESAMIENTO DEL LENGUAJE NATURAL.

Fran Vaquer Estalrich

Quique García Vidal





## **Quique García Vidal**

CEO & Co-founder

*Experto en Inteligencia Artificial. Experiencia en Consultoría de Datos y Orange Spain.*

**Máster en Inteligencia Artificial**

**Máster en Ciencia de Datos**

**Ingeniero Electrónico Industrial**

**(Premio Extraordinario)**

**Research Scientist en IDAL**

<https://www.linkedin.com/in/quique-garcia-vidal/>



## **Fran Vaquer Estalrich**

CTO & Co-founder

*Experto en Inteligencia Artificial. Experiencia como responsable de datos en Aranco.*

**Máster en Big Data**

**Máster en Inteligencia Artificial**

**Ingeniero Electrónico Industrial**

**(Premio Extraordinario)**

**Research Scientist en IDAL**

<https://www.linkedin.com/in/francisco-vaquer-estalrich/>



# deepsense

IA + Investigación + Turismo + Estrategia

## ÍNDICE DE LA SESIÓN.

### 1 INTRODUCCIÓN AL PROCESAMIENTO DEL LENGUAJE NATURAL.

La “explosión” de los datos no estructurados.

Procesamiento del lenguaje natural.

¿Cómo las máquinas entienden y aprenden el lenguaje?

Estado del arte en el procesado del lenguaje natural.

El problema de los modelos del lenguaje.

Transfer Learning.

### 2 [TALLER 1] EXTRACCIÓN DE ENTIDADES PARA ESTRUCTURAR LA INFORMACIÓN.

### 3 [TALLER 2] CREANDO UN CHATBOT CON DIALOGFLOW.

### 4 [TALLER 3] ANÁLISIS DE SENTIMIENTOS CON IA.

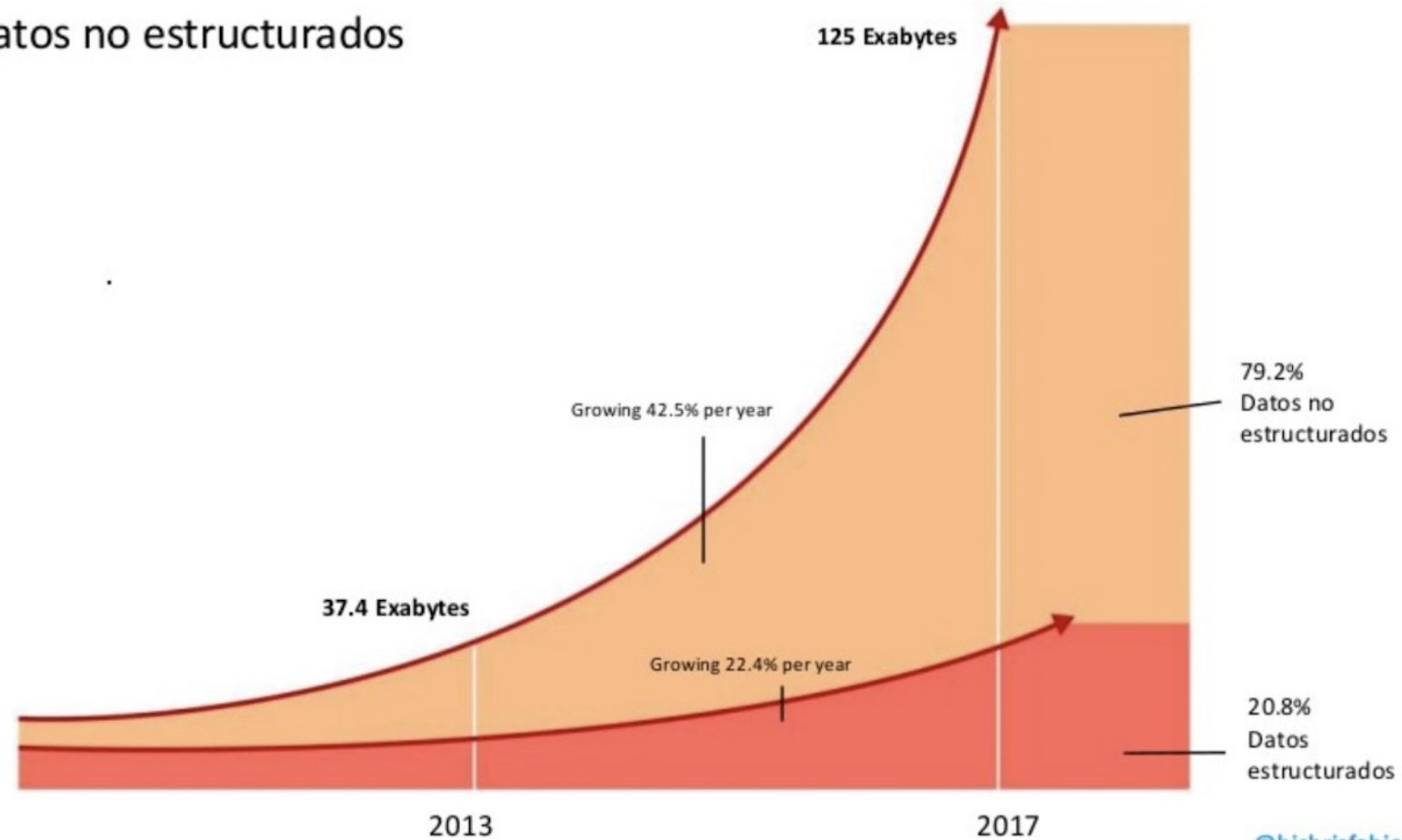


# 1. LA “EXPLOSIÓN” DE LOS DATOS NO ESTRUCTURADOS.

## *Crecimiento en generación de datos*

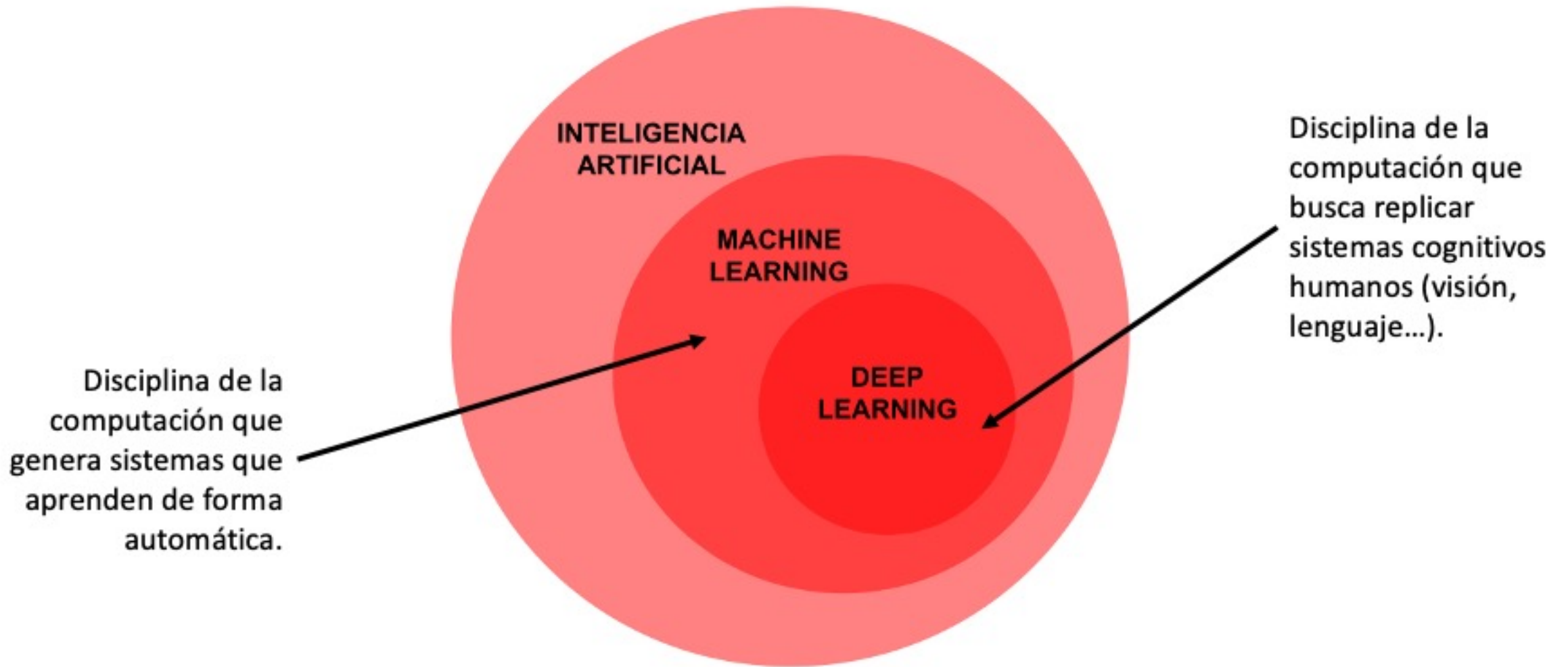
Especialmente datos no estructurados

Source:  
<http://www.slideshare.net/ETCenter/5-onprem-and-the-cloud-effective-big-data-solutions-and-services-for-content-owners>



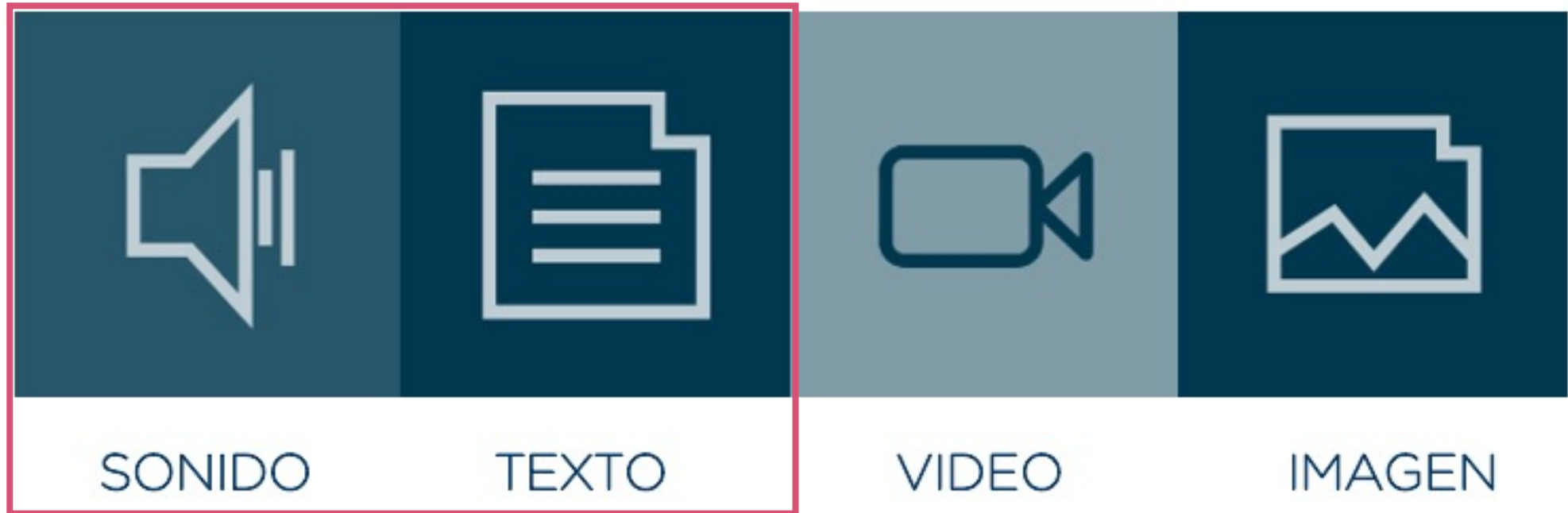
@hichrisfabian

# 1. LA “EXPLOSIÓN” DE LOS DATOS NO ESTRUCTURADOS.



## 1. LA “EXPLOSIÓN” DE LOS DATOS NO ESTRUCTURADOS.

No existen a día de hoy otras tecnologías que nos permitan entender y extraer información de imágenes, texto, sonido y video.



**PROCESAMIENTO DEL LENGUAJE NATURAL**

## 2. PROCESAMIENTO DEL LENGUAJE NATURAL

CHATBOTS

RESÚMENES DE TEXTOS.

TRADUCCIÓN  
AUTOMÁTICA.

EXTRACCIÓN DE  
ENTIDADES.

CLASIFICACIÓN DE  
DOCUMENTOS.

GENERACIÓN DE TEXTO.

ANONIMIZACIÓN DE  
DOCUMENTOS.

ANÁLISIS DE  
SENTIMIENTOS.

...

¿Cómo las “máquinas” entienden el lenguaje?

### 3. ¿CÓMO “LAS MÁQUINAS” ENTIENDEN Y APRENDEN EL LENGUAJE?

#### BAG OF WORDS

	El	perro	es	rojo	come	comida	gato
El perro es rojo. →	1	1	1	1	0	0	0
El perro come comida. →	1	1	0	0	1	1	0
El gato come comida. →	1	0	0	0	1	1	1

---

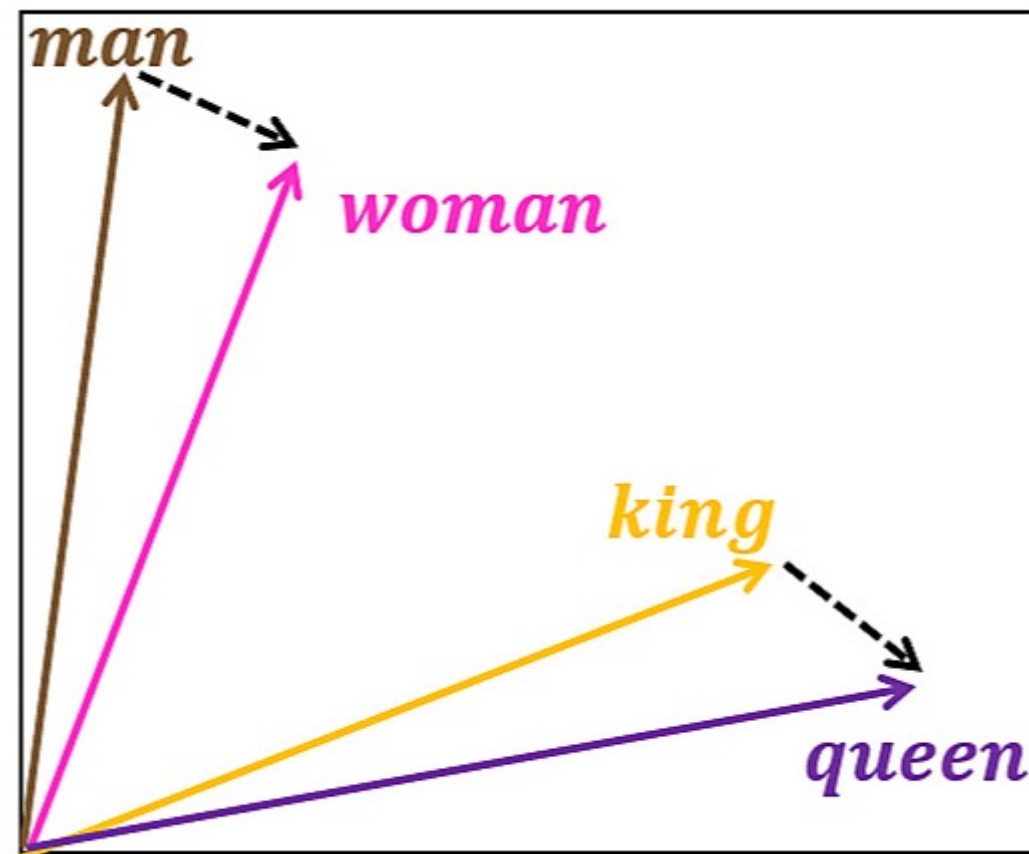
Problemas → {  
  Dimensión (número de columnas = número de palabras).  
  No considera la relación entre las palabras.

### 3. ¿CÓMO “LAS MÁQUINAS” ENTIENDEN Y APRENDEN EL LENGUAJE?

#### EMBEDDINGS:

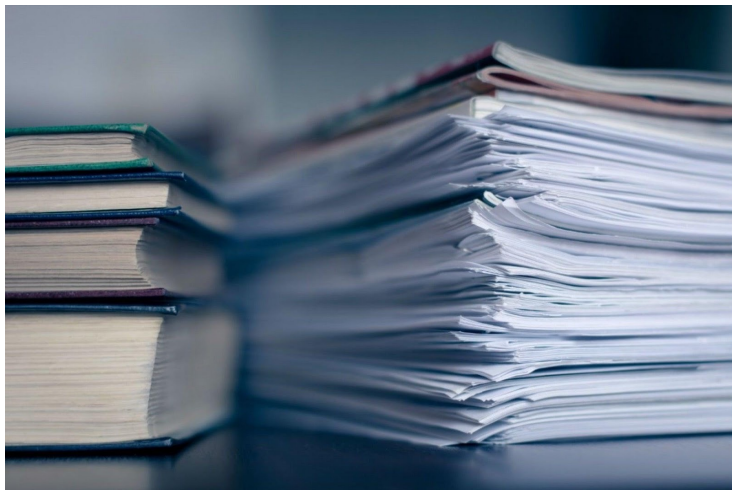
Representaciones numéricas del texto.

<i>man</i>	→	0.6	-0.2	0.8	0.9	-0.1	-0.9	-0.7
<i>woman</i>	→	0.7	0.3	0.9	-0.7	0.1	-0.5	-0.4
<i>king</i>	→	0.5	-0.4	0.7	0.8	0.9	-0.7	-0.6
<i>queen</i>	→	0.8	-0.1	0.8	-0.9	0.8	-0.5	-0.9

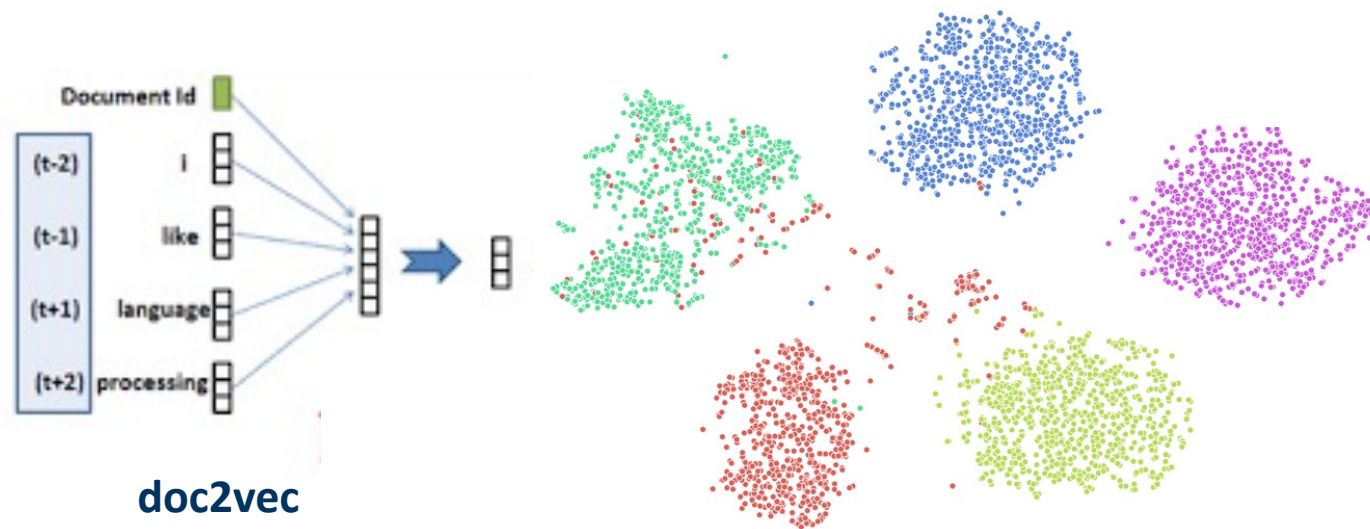


### 3. ¿CÓMO “LAS MÁQUINAS” ENTIENDEN Y APRENDEN EL LENGUAJE?

## DOCUMENT EMBEDDING



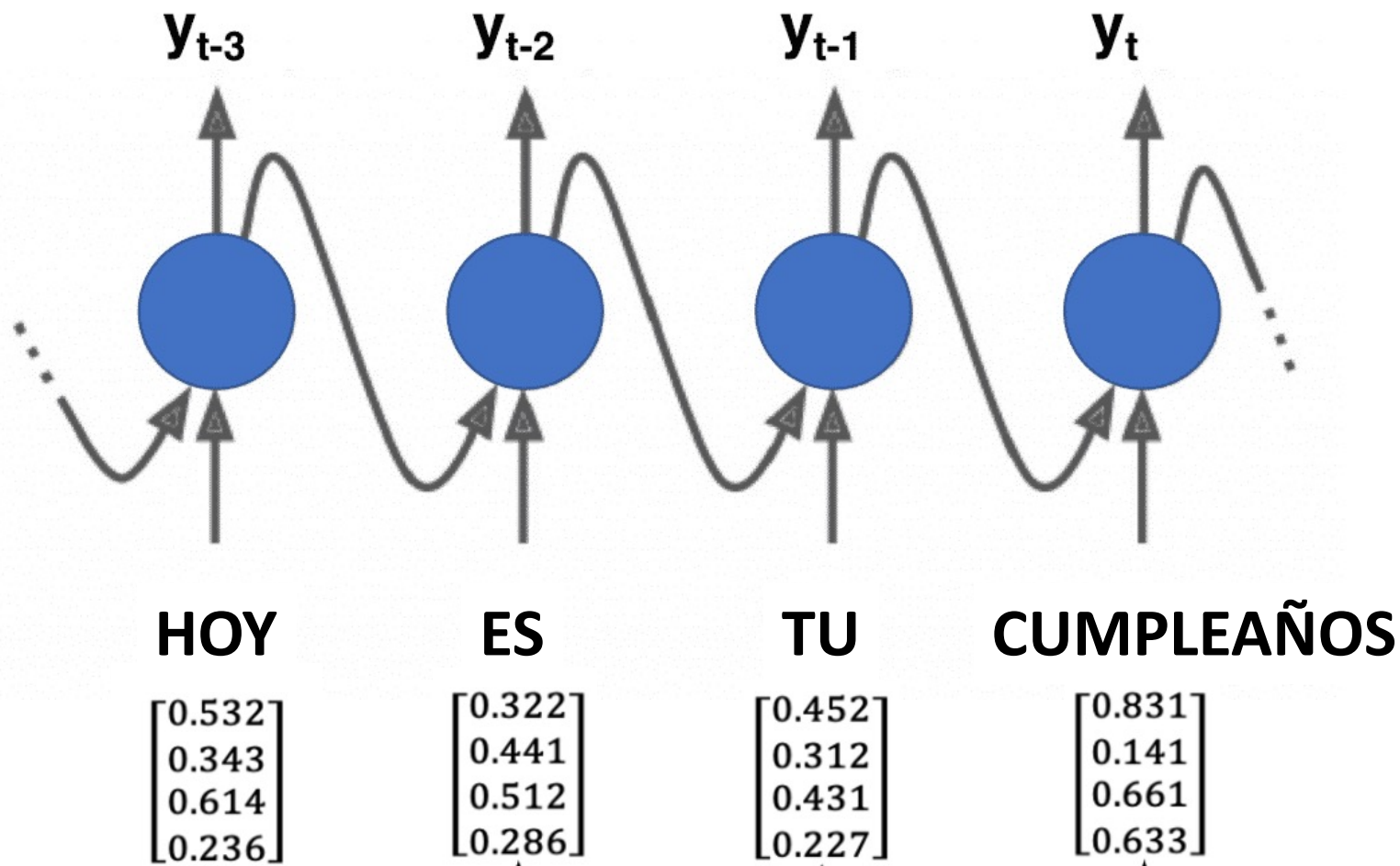
DOCUMENTOS



VECTORES

## 4. ESTADO DEL ARTE EN EL PROCESADO DEL LENGUAJE NATURAL.

### REDES NEURONALES RECURRENTE



### PROBLEMAS DE MEMORIA A LARGO PLAZO

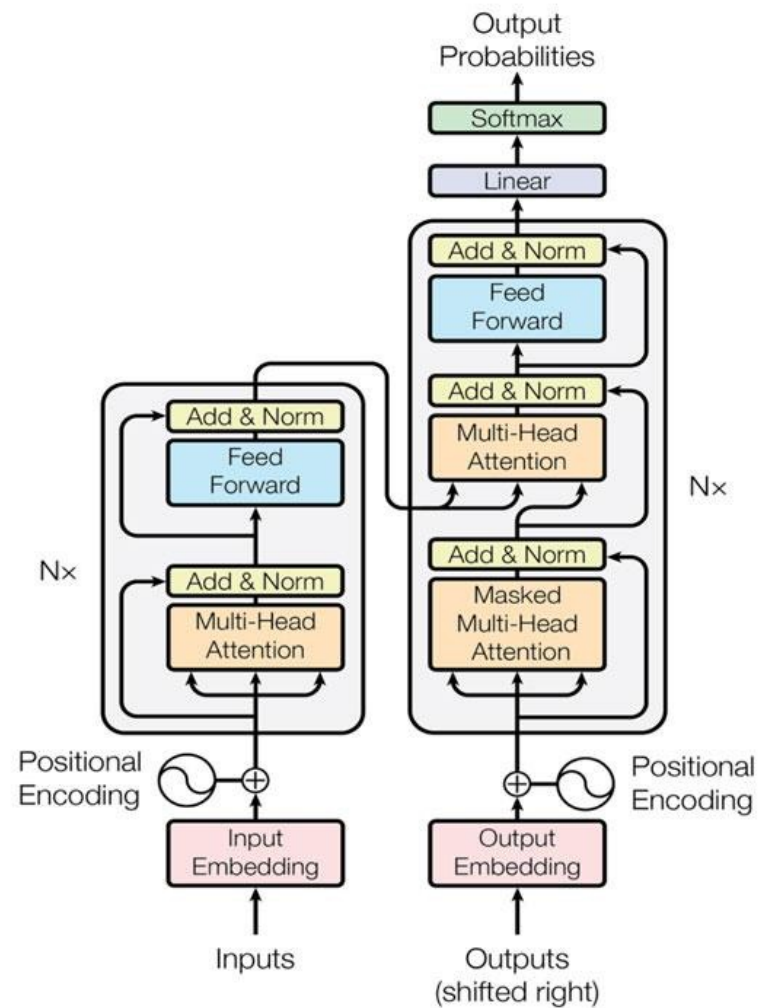
El problema es que cuando la frase es muy larga, el modelo deja de encontrar relaciones con las palabras que han aparecido mucho tiempo atrás

## 4. ESTADO DEL ARTE EN EL PROCESADO DEL LENGUAJE NATURAL.

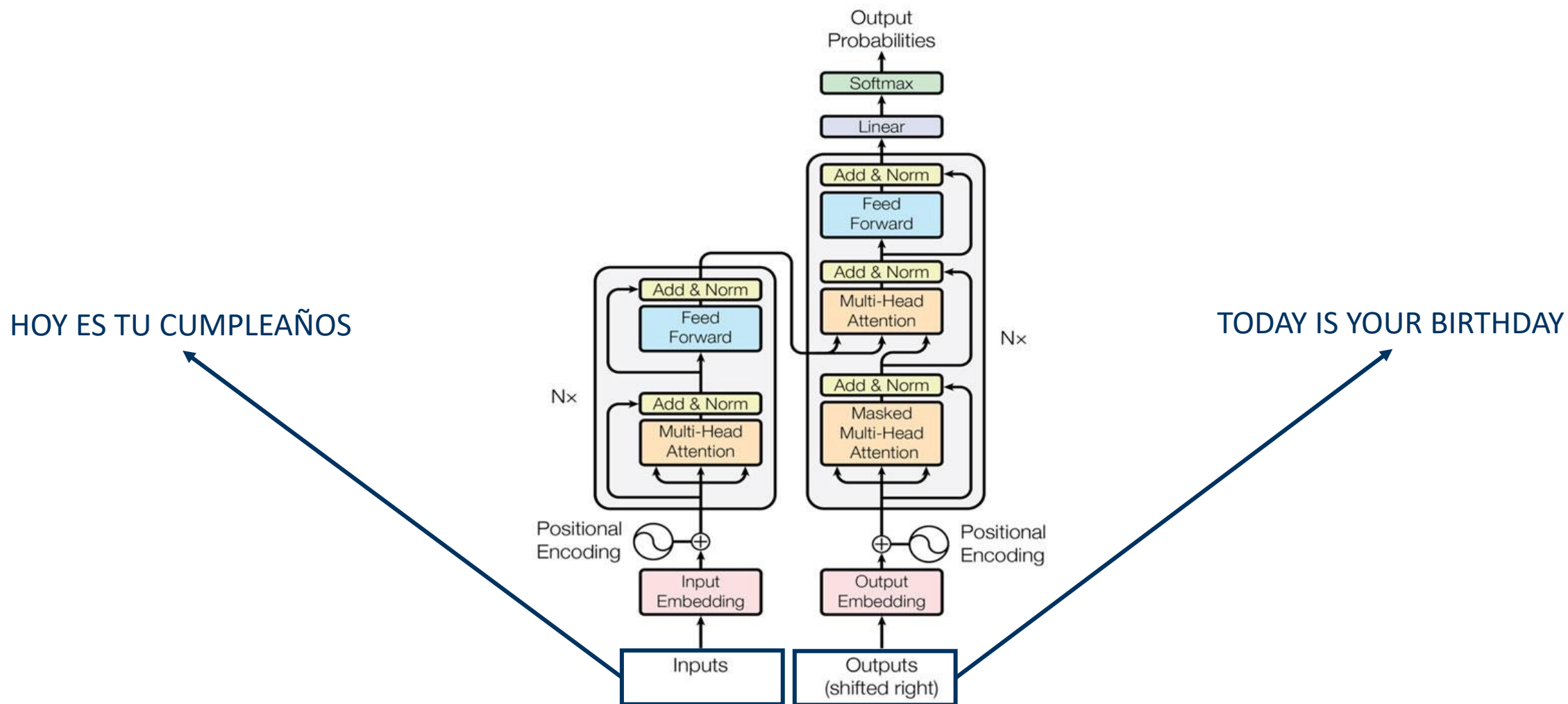
- Estado del arte en tareas de Procesado del Lenguaje Natural: UNA REVOLUCIÓN EN LA IA.

- Attention Is All You Need:

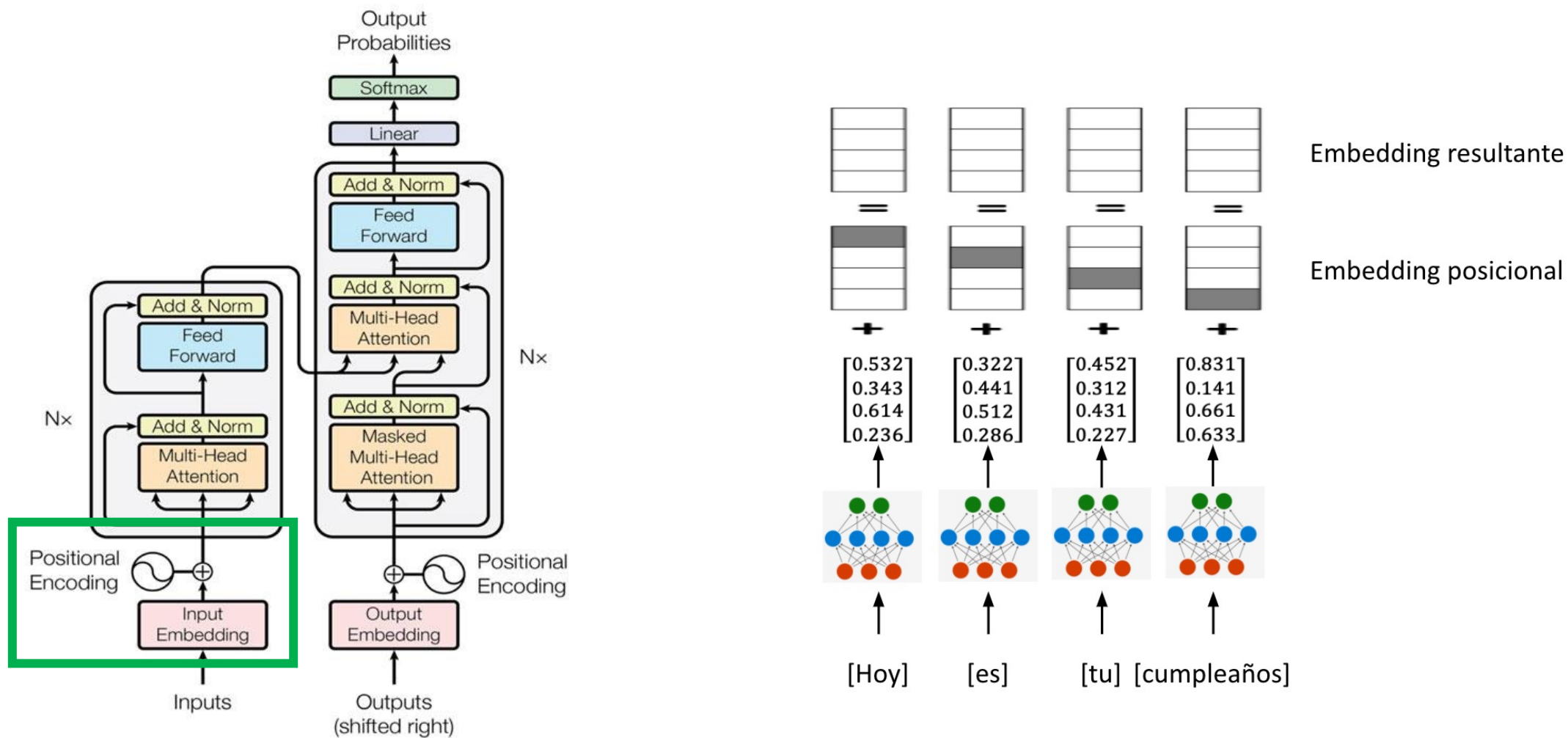
<https://arxiv.org/pdf/1706.03762.pdf>



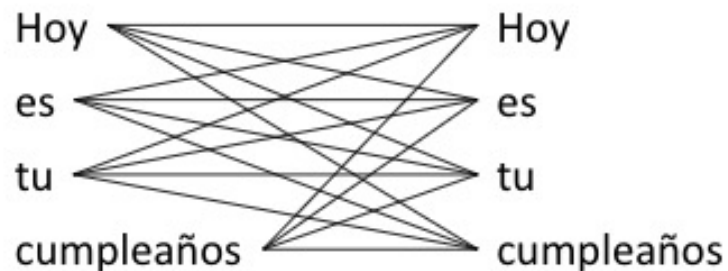
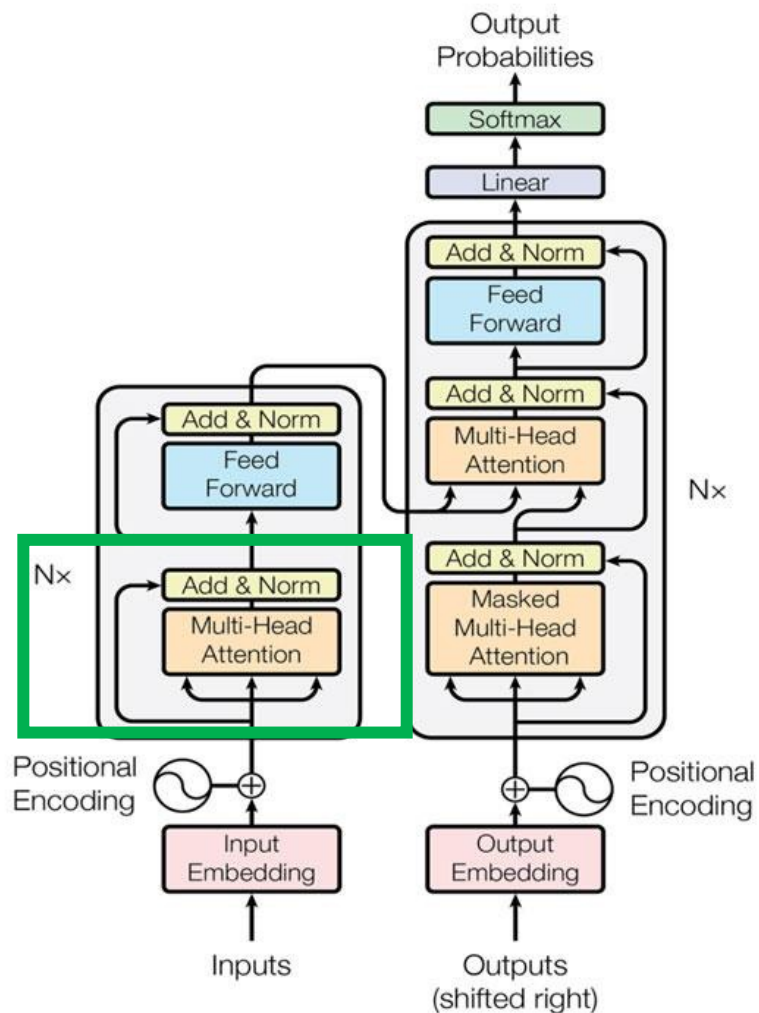
## 4. ESTADO DEL ARTE EN EL PROCESADO DEL LENGUAJE NATURAL.



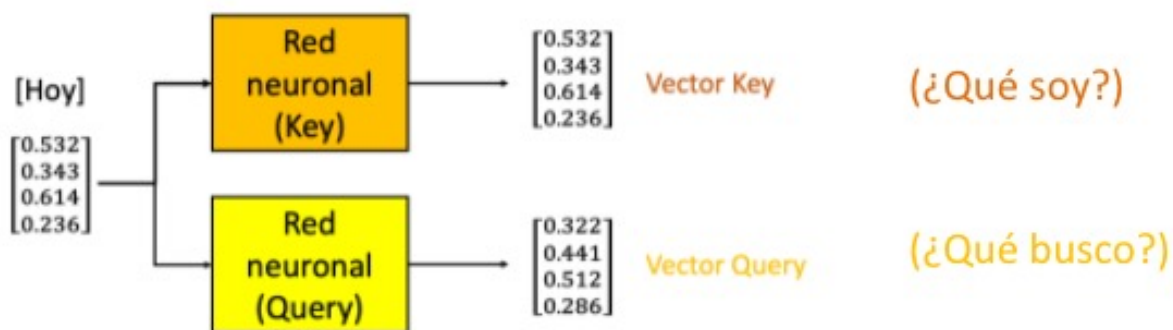
## 4. ESTADO DEL ARTE EN EL PROCESADO DEL LENGUAJE NATURAL.



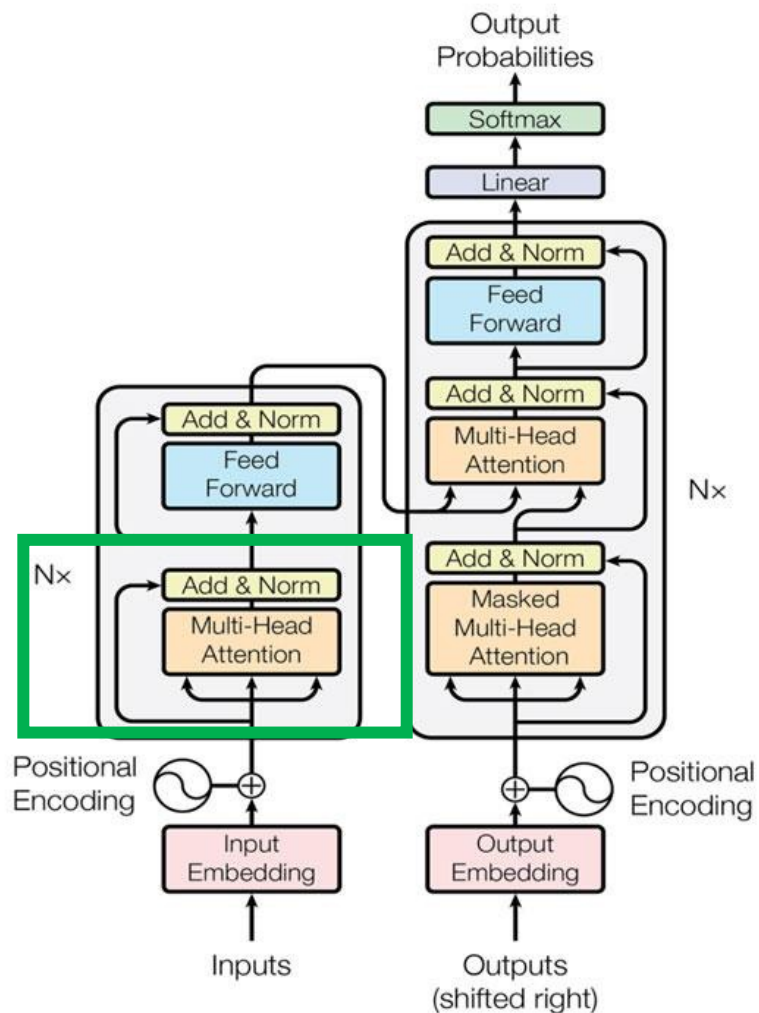
## 4. ESTADO DEL ARTE EN EL PROCESADO DEL LENGUAJE NATURAL.



Buscamos relaciones entre todas las palabras



## 4. ESTADO DEL ARTE EN EL PROCESADO DEL LENGUAJE NATURAL.



### MATRIZ DE ATENCIÓN

	Hoy	es	tu	cumpleaños
Hoy	1	0.2	0.6	0.7
es	0.2	1	0.2	5
tu	0.6	0.2	1	0.9
cumpleaños	0.7	2	0.9	1

¿En qué parte de la frase debe enfocarse cada palabra?

## 4. ESTADO DEL ARTE EN EL PROCESADO DEL LENGUAJE NATURAL.



<https://www.youtube.com/watch?v=D5VN56jQMWM&t=42s>

## 4. ESTADO DEL ARTE EN EL PROCESADO DEL LENGUAJE NATURAL.

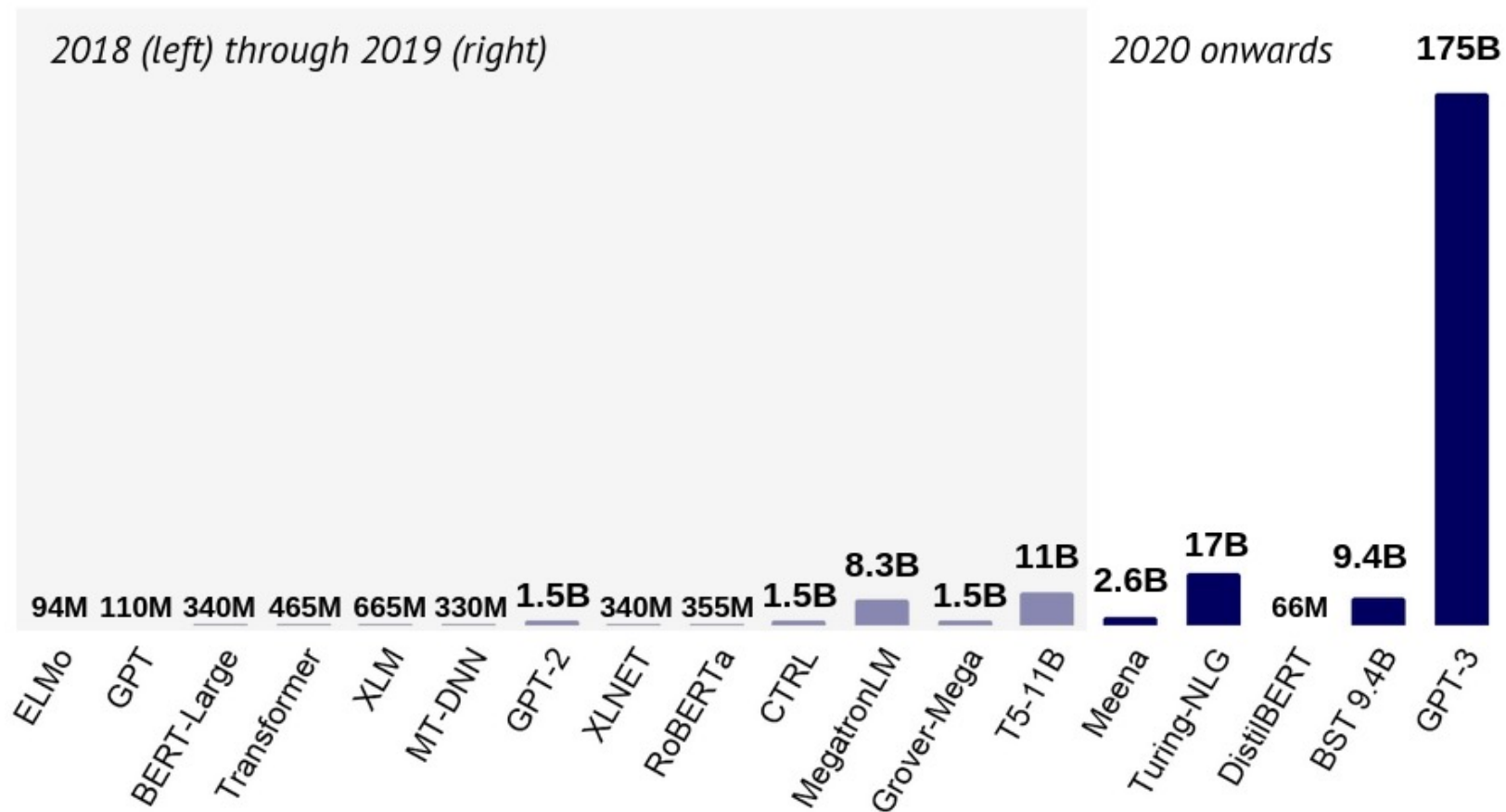


<https://beta.openai.com/playground>

<https://beta.openai.com/examples>

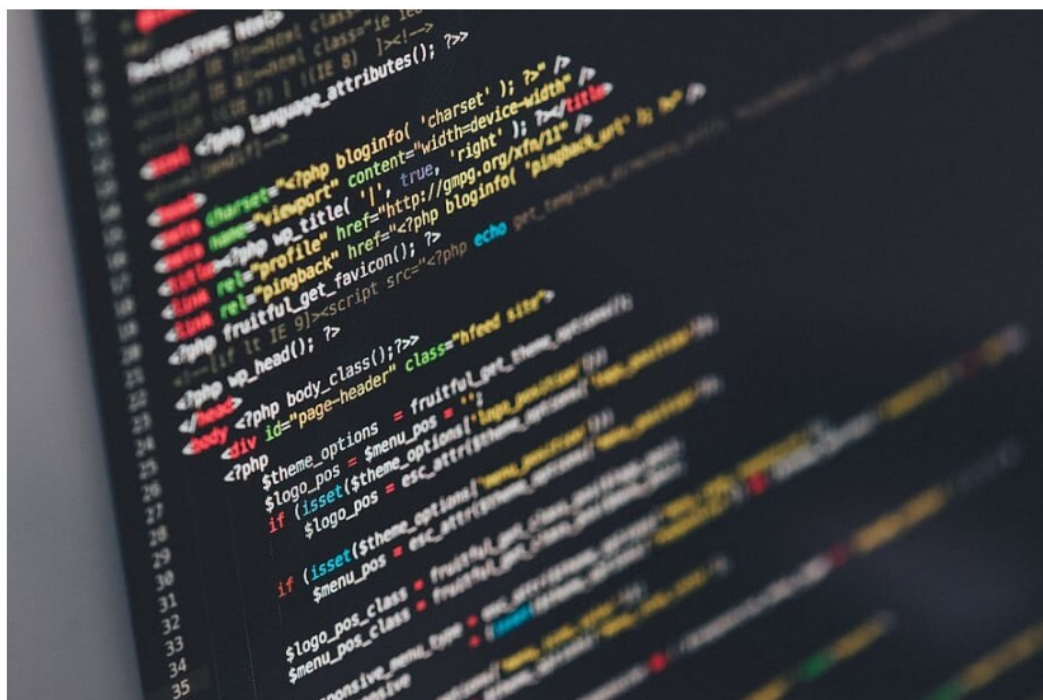
## 5. EL PROBLEMA DE LOS MODELOS DEL LENGUAJE.

### EVOLUCIÓN DE MODELOS DEL LENGUAJE

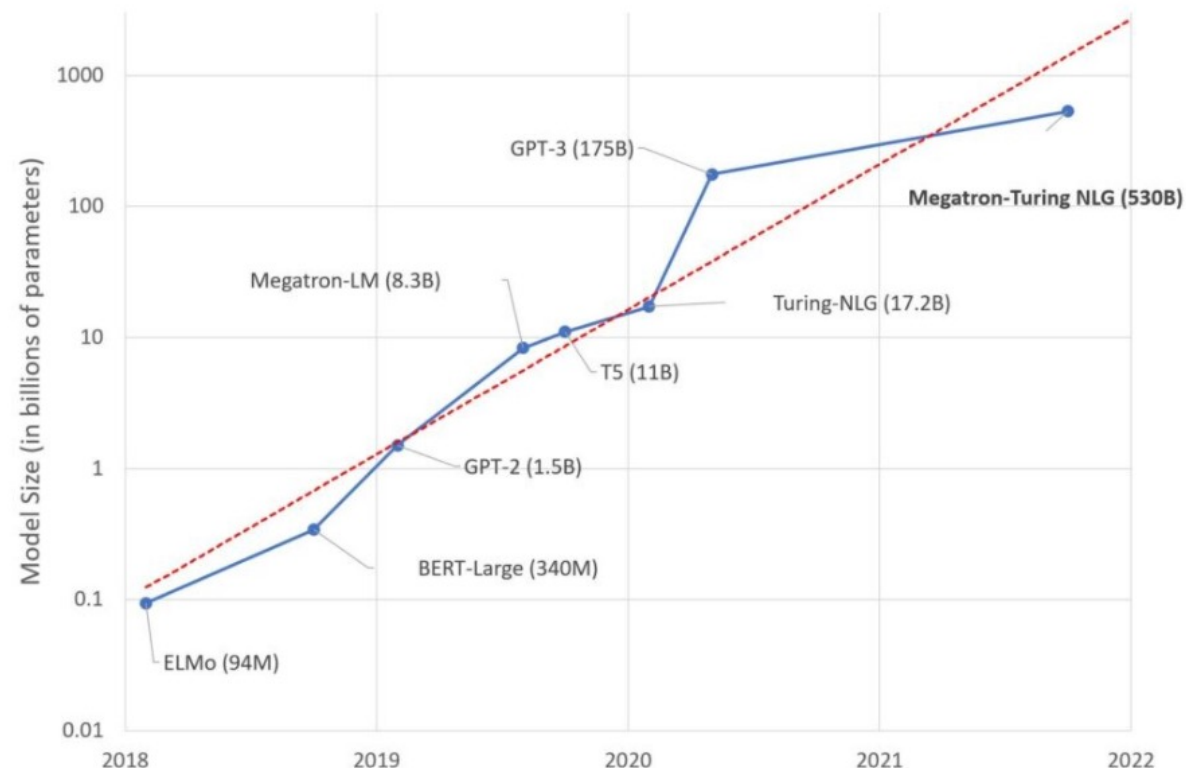


## 5. EL PROBLEMA DE LOS MODELOS DEL LENGUAJE.

Microsoft y Nvidia lanzan Megatron-Turing, "el modelo de lenguaje más potente del mundo", con el triple de parámetros que GPT-3



5 Comentarios



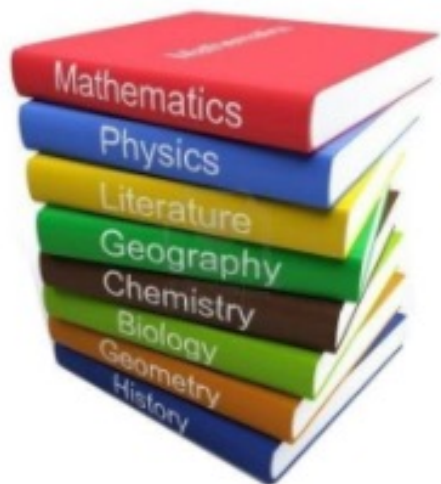
13 Octubre 2021 - Actualizado 13 Octubre 2021, 10:46

## 5. EL PROBLEMA DE LOS MODELOS DEL LENGUAJE.

Suponen un coste computacional (y económico) y un impacto medioambiental enorme y poco accesible para el usuario medio.



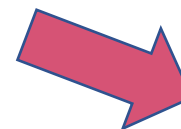
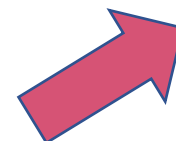
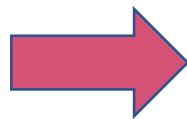
## 6. TRANSFER LEARNING: LA DEMOCRATIZACIÓN DEL LENGUAJE.



APRENDE CONOCIMIENTO GENERAL

ESPECIALIZACIÓN  
(vocabulario, temas, tareas...)

## 6. TRANSFER LEARNING: LA DEMOCRATIZACIÓN DEL LENGUAJE.



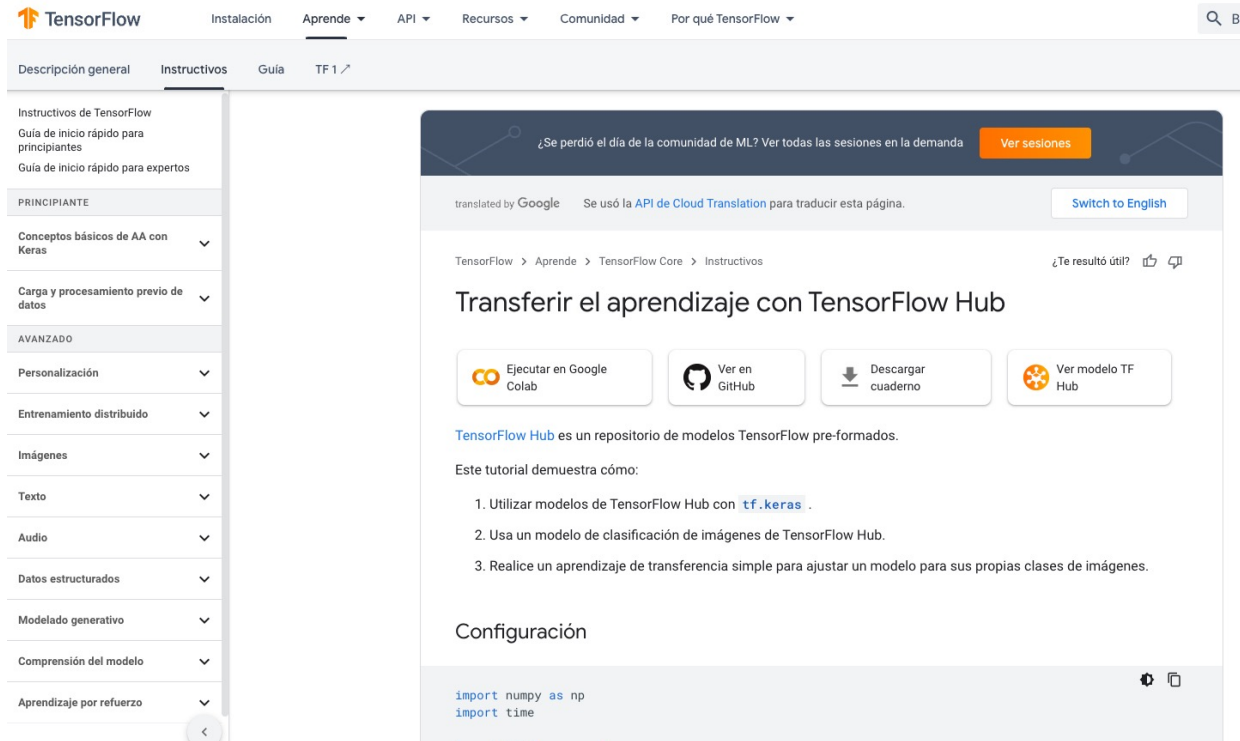
CAMPOS ESPECÍFICOS  
(salud, Agro, Industria, legal...)

TAREAS ESPECÍFICAS  
(Traducción, extracción  
entidades, clasificación,  
análisis de sentimiento...)

APRENDE CONOCIMIENTO GENERAL


ESPECIALIZACIÓN  
"Fine Tuning"

## 6. TRANSFER LEARNING: LA DEMOCRATIZACIÓN DEL LENGUAJE.



The screenshot shows the TensorFlow Hub website interface. The top navigation bar includes 'TensorFlow', 'Instalación', 'Aprende', 'API', 'Recursos', 'Comunidad', and 'Por qué TensorFlow'. The main content area is titled 'Transferir el aprendizaje con TensorFlow Hub' and includes a 'Ver sesiones' button. Below the title, there are four action buttons: 'Ejecutar en Google Colab', 'Ver en GitHub', 'Descargar cuaderno', and 'Ver modelo TF Hub'. The text explains that TensorFlow Hub is a repository of pre-trained models and provides a list of steps for using transfer learning. A 'Configuración' section shows code snippets for importing numpy and time.

<https://www.tensorflow.org/hub>











**The AI community  
building the future.**

Build, train and deploy state of the art models powered by  
the reference open source in machine learning.

Star 59,370

More than 5,000 organizations are using Hugging Face

 <b>Allen Institute for AI</b> Non-Profit - 83 models	 <b>Facebook AI</b> Company - 248 models	 <b>Graphcore</b> Company - 16 models	 <b>Google AI</b> Company - 492 models
 <b>Amazon Web Services</b> Company - 1 model	 <b>SpeechBrain</b> Non-Profit - 36 models	 <b>Microsoft</b> Company - 140 models	 <b>Grammarly</b> Company

<https://huggingface.co/>

## 6. TRANSFER LEARNING: LA DEMOCRATIZACIÓN DEL LENGUAJE.

### EJEMPLOS HUGGING FACE



**The AI community  
building the future.**

Build, train and deploy state of the art models powered by  
the reference open source in machine learning.

Star 59,370

More than 5,000 organizations are using Hugging Face

<b>Allen Institute for AI</b> Non-Profit · 83 models	<b>Facebook AI</b> Company · 248 models	<b>Graphcore</b> Company · 16 models	<b>Google AI</b> Company · 492 models
<b>Amazon Web Services</b> Company · 1 model	<b>SpeechBrain</b> Non-Profit · 36 models	<b>Microsoft</b> Company · 140 models	<b>Grammarly</b> Company

<https://huggingface.co/>

**Question-Answering**  
(Responder preguntas)

<https://huggingface.co/tasks/question-answering>

**Resúmenes de textos**

[https://huggingface.co/mrm8488/bert2bert\\_shared-spanish-finetuned-summarization](https://huggingface.co/mrm8488/bert2bert_shared-spanish-finetuned-summarization)

**Extracción de entidades**  
(Nombres, ciudades, organizaciones...)

<https://huggingface.co/Davlan/bert-base-multilingual-cased-ner-hrl>

## 6. TRANSFER LEARNING: LA DEMOCRATIZACIÓN DEL LENGUAJE.

### Analítica de textos “clásica”



### Deep Learning (Inteligencia Artificial). Transfer Learning



**[TALLER 1]**

**EXTRACCIÓN DE ENTIDADES PARA ESTRUCTURAR LA INFORMACIÓN.**

# [TALLER 1] EXTRACCIÓN DE ENTIDADES PARA ESTRUCTURAR LA INFORMACIÓN.

La extracción de entidades nos permite pasar de información no estructurada a información estructurada y automatizar procesos de extracción de información.

## INPUT MODELO

**España propone a Bruselas limitar el gas a 30 euros para que la luz baje a la mitad**

La propuesta, que aún necesita el visto bueno de la Comisión Europea, llevaría el coste medio diario de la electricidad hasta los 120 o 130 euros por megavatio hora. La idea sería mantener ese tope hasta final de año



## OUTPUT MODELO

España **LOC** propone a Bruselas **LOC** limitar el gas a 30 euros para que la luz baje a la mitadLa propuesta, que aún necesita el visto bueno de la Comisión Europea **ORG** , llevaría el coste medio diario de la electricidad hasta los 120 o 130 euros por megavatio hora. La idea sería mantener ese tope hasta final de año

<https://elpais.com/economia/2022-03-31/espana-y-portugal-proponen-a-bruselas-limitar-a-30-euros-el-precio-del-gas-para-las-centrales-termicas.html>

**[TALLER 2]**  
**CREANDO UN CHATBOT CON DIALOGFLOW.**

**[TALLER 3]**  
**ANÁLISIS DE SENTIMIENTOS CON IA.**



# deepsense

An Artificial Intelligence Company.

[www.deepsense.es](http://www.deepsense.es)  
[contacto@deepsense.es](mailto:contacto@deepsense.es)