

The background features a dark blue world map composed of small dots, overlaid with faint binary code (0s and 1s) and abstract geometric lines.

FAKE

NEWS

CON GPT-2 COMO INVITADO ESPECIAL

DISCLAIMER

El objetivo de la presentación es hacer una introducción a los modelos de generación de texto, en concreto el GPT-2, y demostrar su uso.

En ningún caso está orientada a la generación o divulgación de información falsa.

¿Qué vamos a contar?

Introducción a los modelos de generación de texto

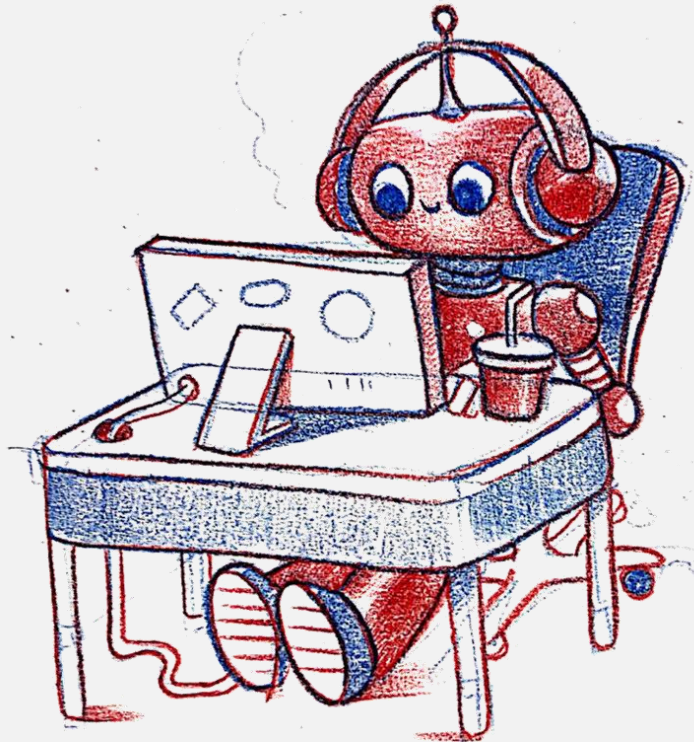
- Evolución de los modelos de generación de texto
- Introducción a la generación de texto con GPT-2 en Huggingface

Como controlar la temática y la estructura del texto que se genera

- Fine-tuning
- Plug and Play Language Model

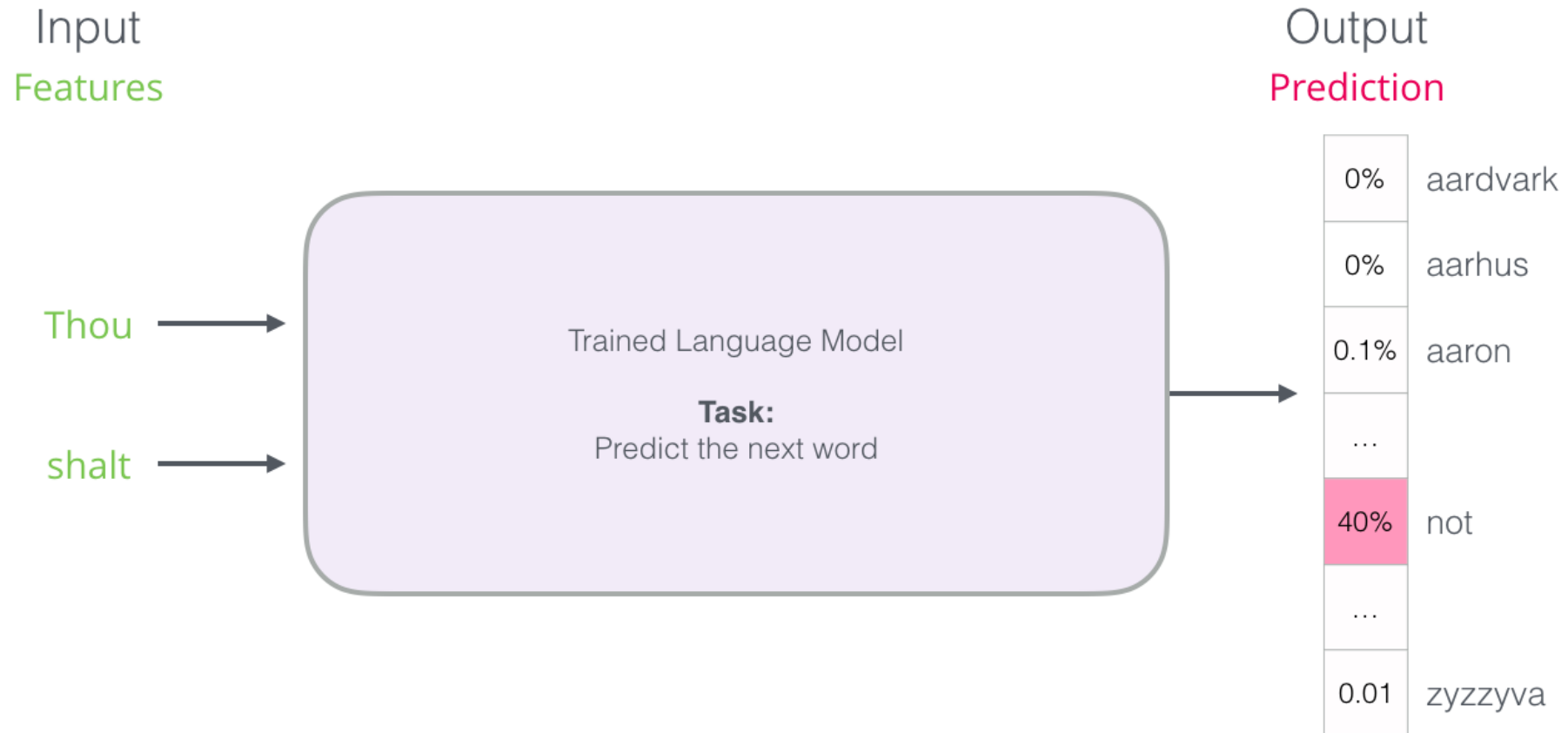
Reflexión sobre el uso ético de estos modelos

MODELOS DE GENERACIÓN DE TEXTO



¿Qué es un modelo de generación de texto?

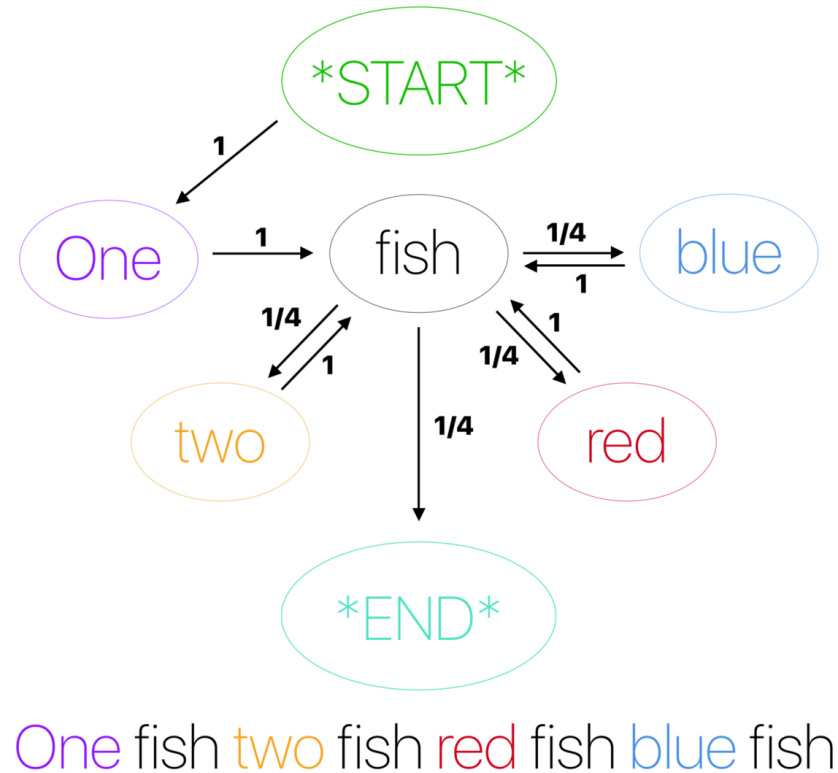
Es un modelo capaz de, dado un texto, predecir la siguiente palabra o secuencia de palabras



Evolución de los modelos de generación de texto

1. Cadenas de Markov

- Predicen una palabra a partir de la anterior
- Cada palabra es un estado en la cadena
- La probabilidad de la siguiente palabra (en base a la anterior) se calcula en base al número de ocurrencias de ambas palabras consecutivamente en el texto de entrenamiento
- Problemas:
 - Al focalizarse sólo en la última palabra pierden el contexto de todo lo anterior
 - Al aumentar el vocabulario y la longitud del texto se vuelven demasiado grandes y complejas



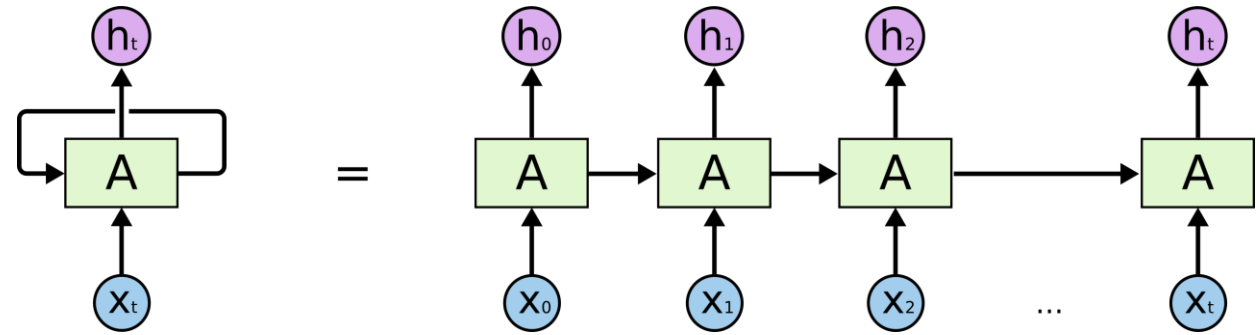
Fuente: [hackernoon](#)

Evolución de los modelos de generación de texto

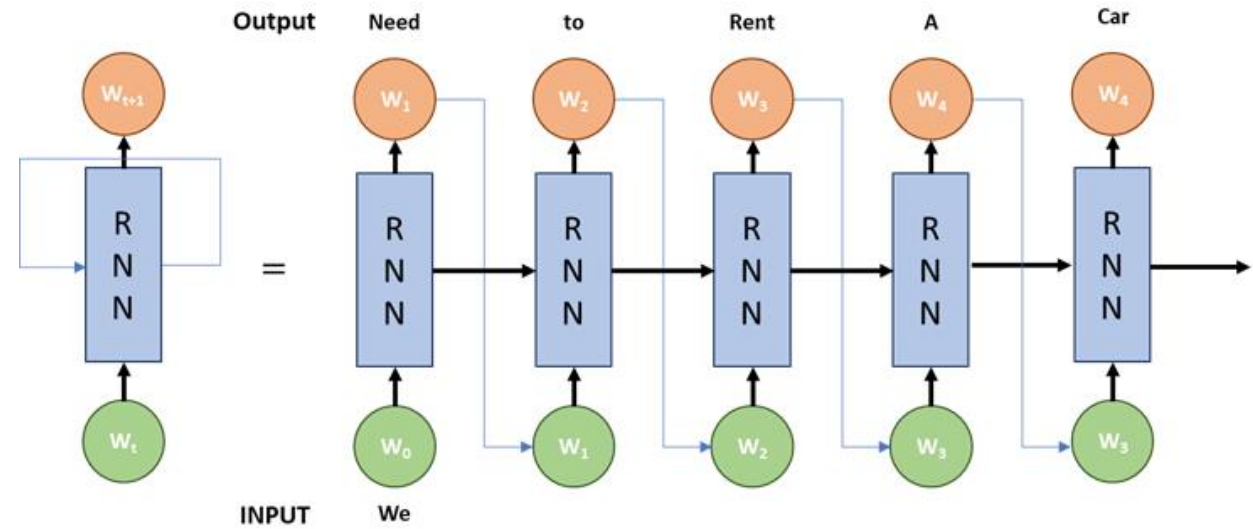
1. Cadenas de Markov

2. RNN (Redes Neuronales Recurrentes)

- Se usan para datos secuenciales, como el texto.
- Son un tipo de red neuronal con bucles, permitiendo que la información se persista durante más tiempo.
- A diferencia de las cadenas de Markov, retienen el contexto del texto anterior a la hora de hacer predicciones
- Problemas:
 - Problema de desvanecimiento de gradiente: para secuencias de texto muy largas, las RNNs “olvidan” palabras que vieron muy atrás, limitando la longitud del texto que pueden generar.



Fuente: [github](#)



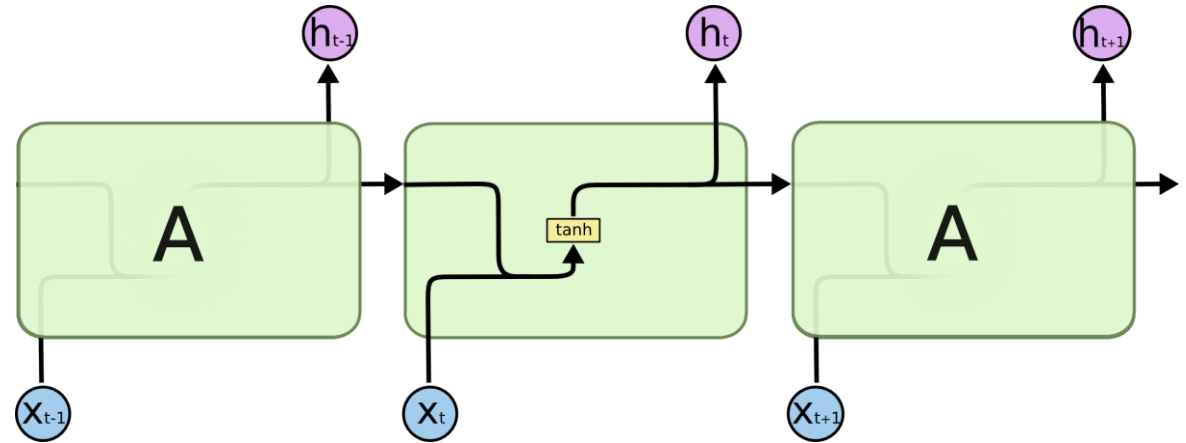
Evolución de los modelos de generación de texto

1. Cadenas de Markov

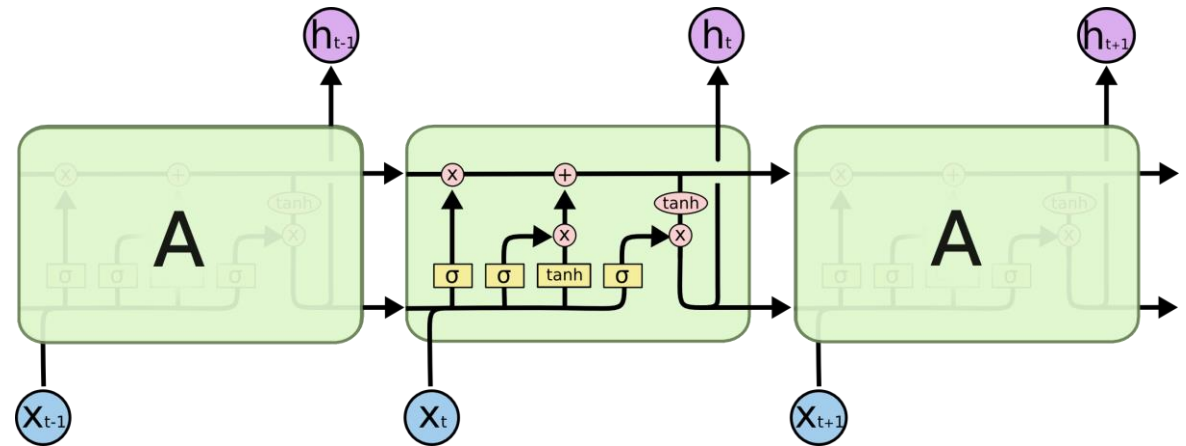
2. RNN (Redes Neuronales Recurrentes)

3. Long Short-Term Memory (LSTM)

- Son un tipo de RNN capaces de aprender dependencias con estados anteriores y utilizarlos para la predicción. Es decir, que tienen “memoria” a más largo plazo.
- Esto se consigue mediante una estructura más compleja en la que cada neurona tiene tres puertas: entrada, olvido y salida.
- Problemas:
 - Siguen estando limitadas en cuanto a lo que pueden recordar.
 - Difíciles de entrenar:
 - Requieren de muchos recursos
 - Son difíciles de paralelizar



Fuente: [github](#)



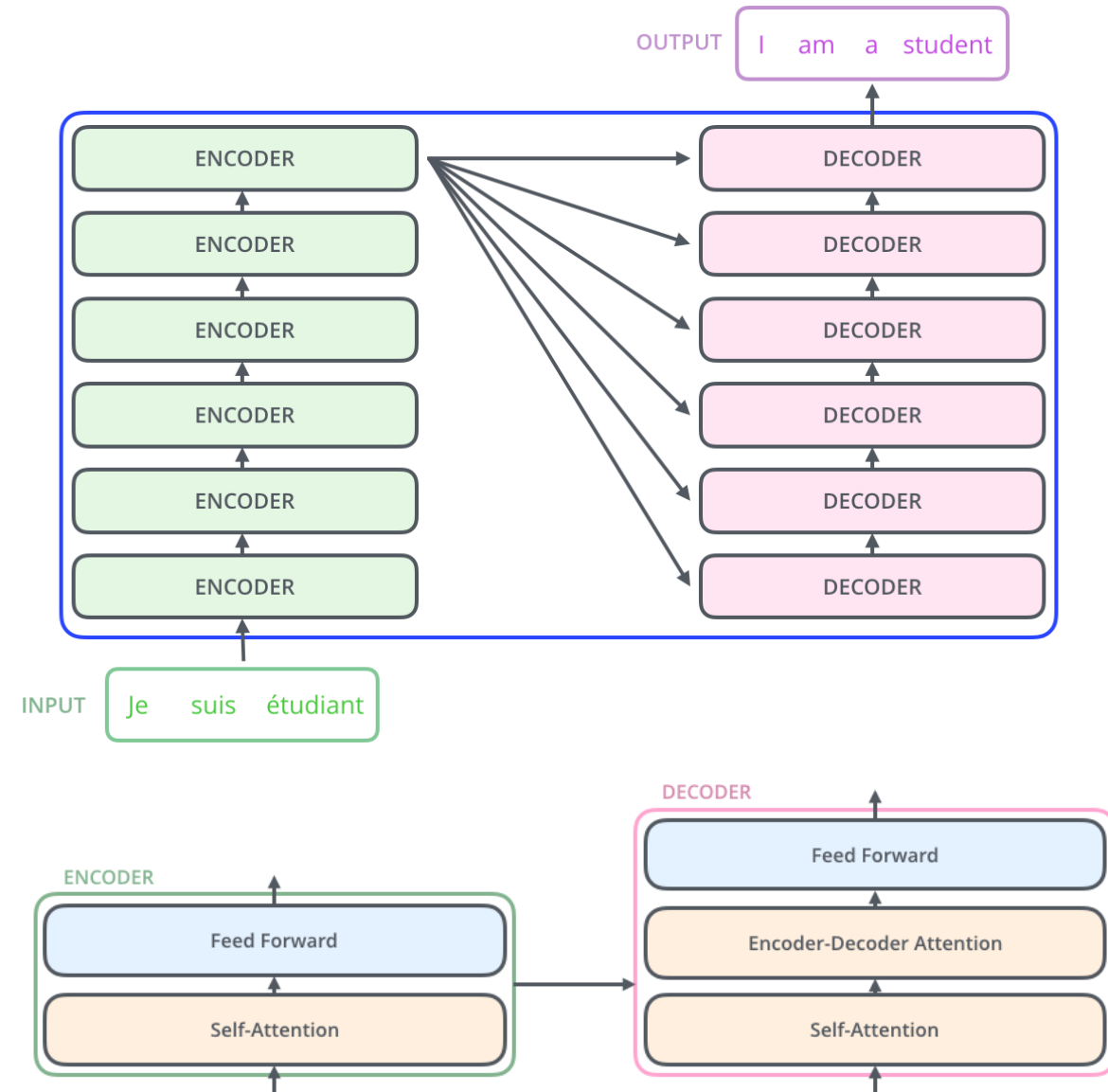
Fuente: [github](#)

Evolución de los modelos de generación de texto

1. Cadenas de Markov
2. RNN (Redes Neuronales Recurrentes)
3. Long Short-Term Memory (LSTM)

4. Transformer

- Es una arquitectura propuesta en el paper de Google de 2017 “[Attention Is All You Need](#)”
- Las principales innovaciones que introduce son:
 - Capas de atención: permiten aprender que palabras son las más relevantes, es decir, codifican el contexto del texto para cada palabra
 - Positional embedding: añade información de la posición de cada palabra dentro de la secuencia.
 - Se puede paralelizar.

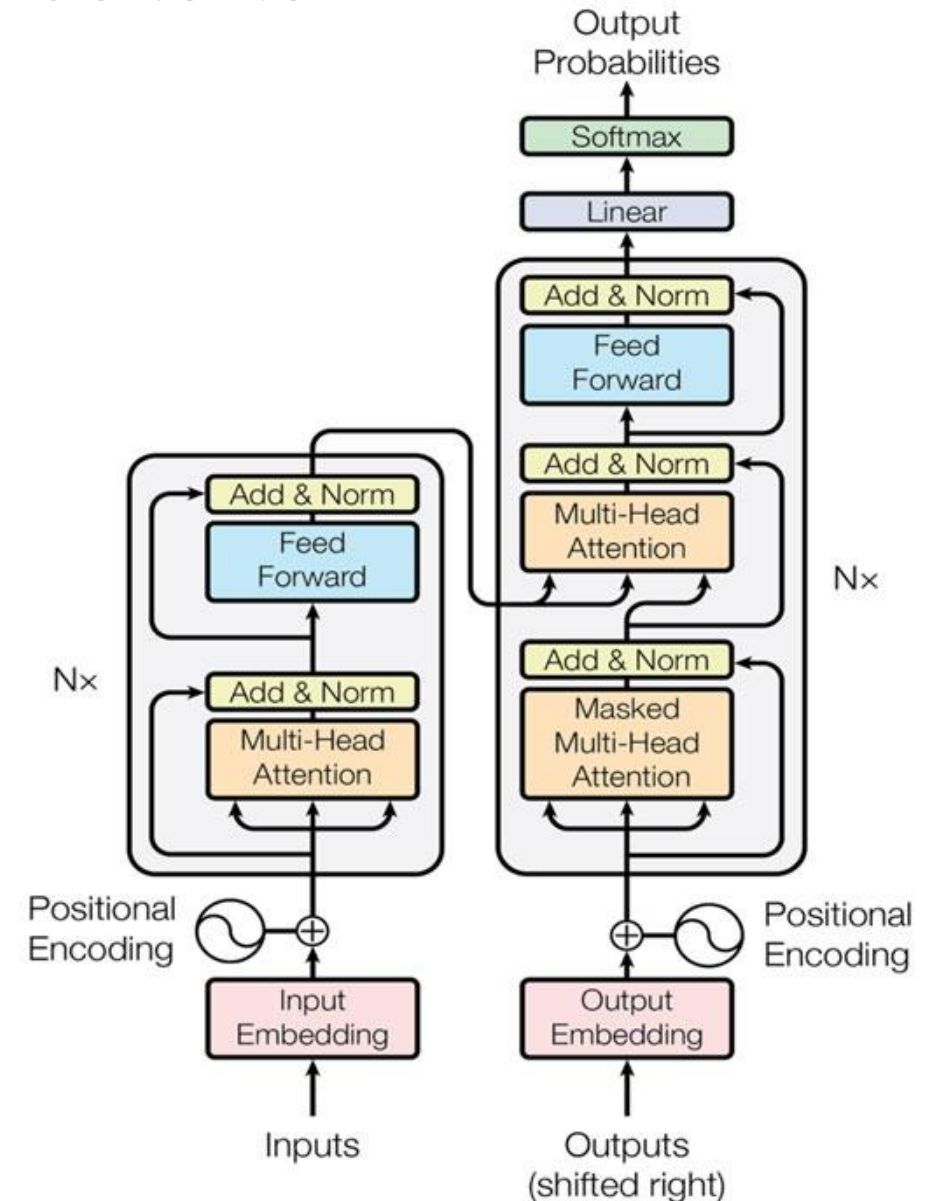


Evolución de los modelos de generación de texto

1. Cadenas de Markov
2. RNN (Redes Neuronales Recurrentes)
3. Long Short-Term Memory (LSTM)

4. Transformer

- Es una arquitectura propuesta en el paper de Google de 2017 "[Attention Is All You Need](#)"
- Las principales innovaciones que introduce son:
 - Capas de atención: permiten aprender que palabras son las más relevantes, es decir, codifican el contexto del texto para cada palabra
 - Positional embedding: añade información de la posición de cada palabra dentro de la secuencia.
 - Se puede paralelizar.

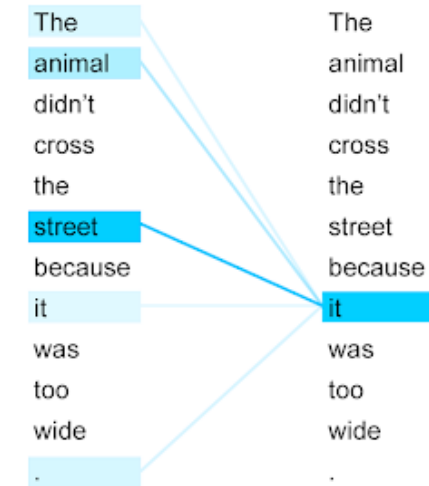
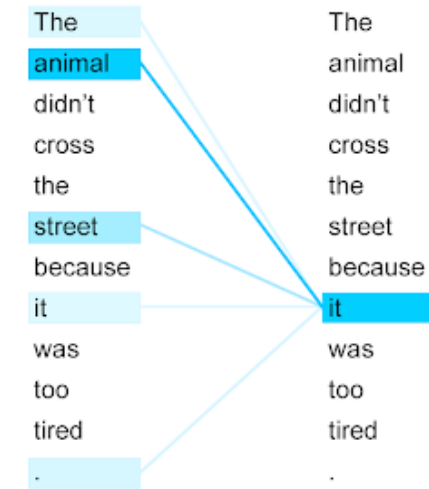


Evolución de los modelos de generación de texto

1. Cadenas de Markov
2. RNN (Redes Neuronales Recurrentes)
3. Long Short-Term Memory (LSTM)

4. Transformer

- Es una arquitectura propuesta en el paper de Google de 2017 "[Attention Is All You Need](#)"
- Las principales innovaciones que introduce son:
 - **Capas de atención:** permiten aprender que palabras son las más relevantes, es decir, codifican el contexto del texto para cada palabra
 - Positional embedding: añade información de la posición de cada palabra dentro de la secuencia.
 - Se puede paralelizar.

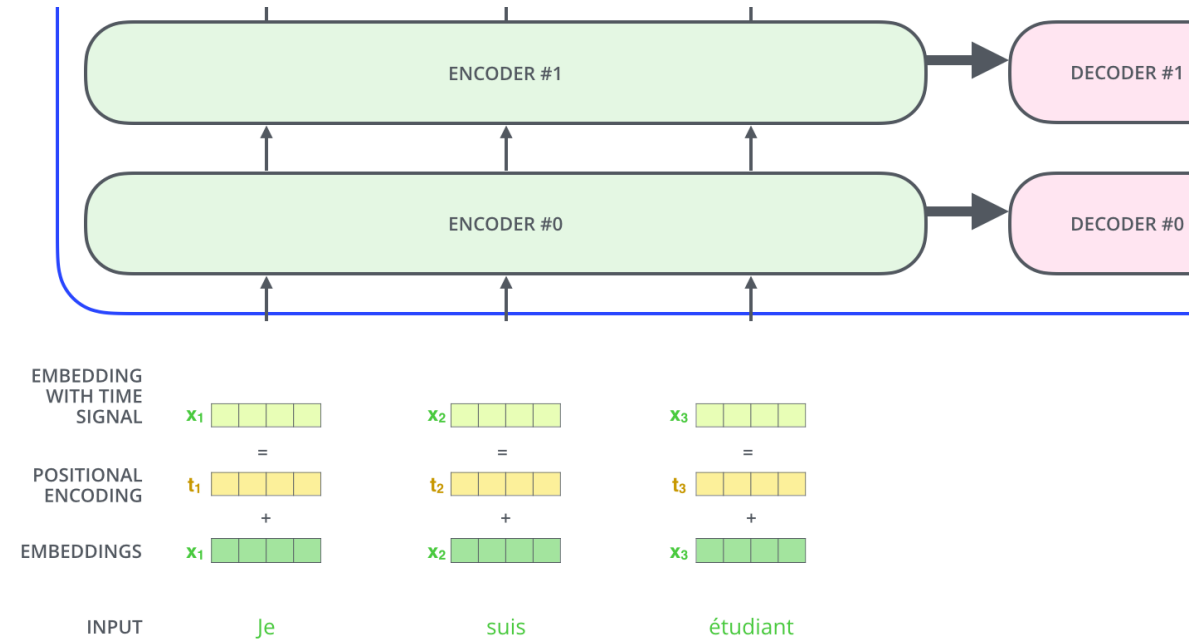


Evolución de los modelos de generación de texto

1. Cadenas de Markov
2. RNN (Redes Neuronales Recurrentes)
3. Long Short-Term Memory (LSTM)

4. Transformer

- Es una arquitectura propuesta en el paper de Google de 2017 "[Attention Is All You Need](#)"
- Las principales innovaciones que introduce son:
 - Capas de atención: permiten aprender que palabras son las más relevantes, es decir, codifican el contexto del texto para cada palabra
 - **Positional embedding**: añade información de la posición de cada palabra dentro de la secuencia.
 - Se puede paralelizar.



Fuente: [googleblog](#)

Modelos basados en Transformer



GPT-2

DECODER

...

DECODER

DECODER



BERT

ENCODER

...

ENCODER

ENCODER



TRANSFORMER XL

RECURRENT DECODER

...

RECURRENT DECODER

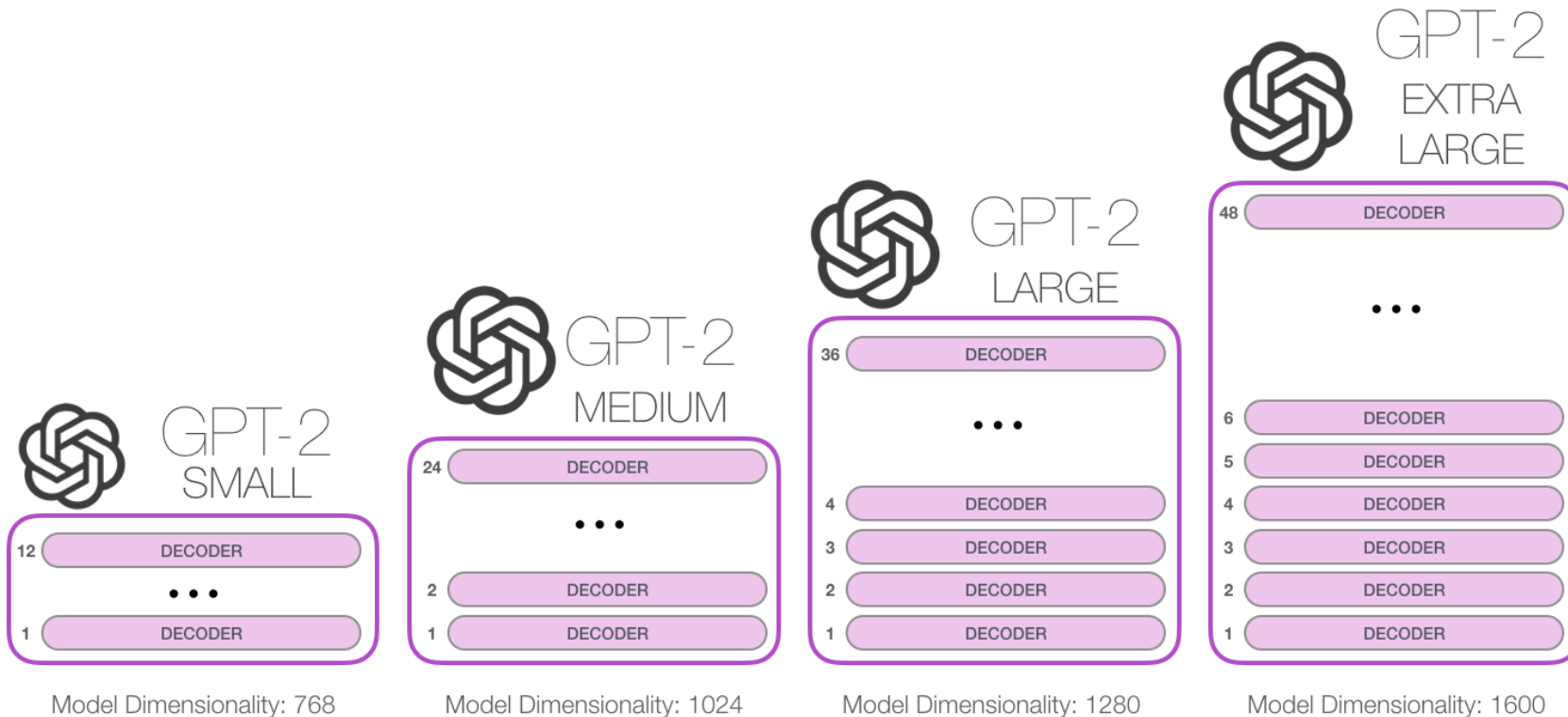
RECURRENT DECODER

OpenAI GPT-2

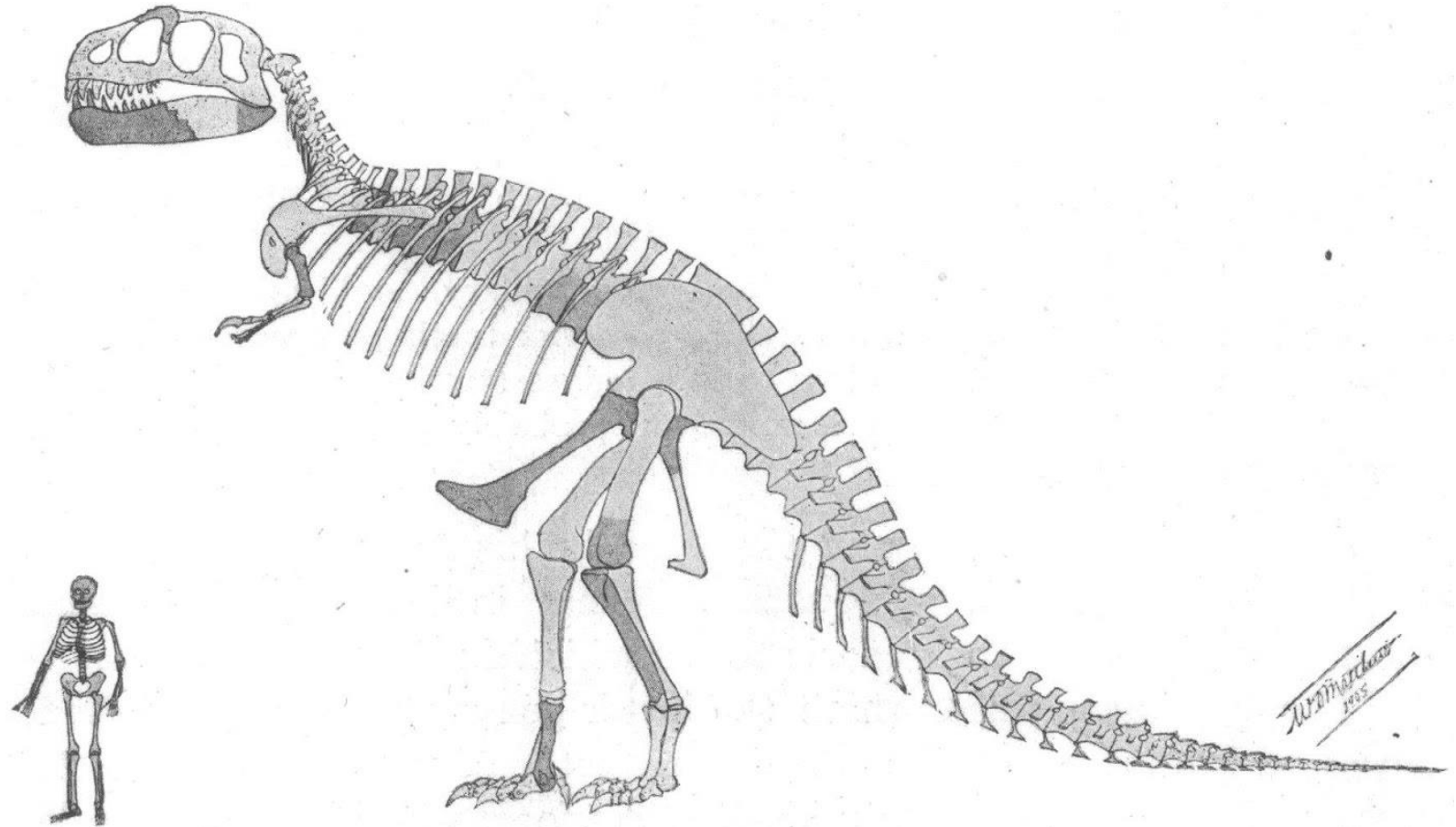


Es un modelo de generación de texto creado por OpenAI en febrero de 2019 basado en la arquitectura del Transformer.

- Es autorregresivo, es decir, genera un token en cada iteración
- Está disponible en varios tamaños dependiendo del tamaño del embedding



OpenAI GPT-3



GPT-2
1.5B Parameters

GPT-3
175B Parameters

Huggingface Transformers

Librería de Python que descarga modelos pre-entrenados para tareas de:

- Comprensión del lenguaje natural, como análisis de sentimiento
- Generación de lenguaje natural como generación de texto o traducción de texto



Conceptos o clases principales

- **Tokenizer:** almacenan el vocabulario de cada modelo e incluye métodos para codificar y decodificar string en una lista de índices token embeddings que sirven de entrada al modelo
- **Configuration:** contienen los parámetros necesarios para construir un modelo. No es necesaria cuando se usa un modelo pre-entrenado
- **Model:** modelos de [Pytorch](#) o Keras para trabajar con los modelos pre-entrenados por la librería.



Huggingface Transformers - Generación texto con GPT-2

A la hora de generar texto, huggingface dispone de varios métodos de decodificación (o formas de elegir la siguiente palabra o palabras dada una frase).

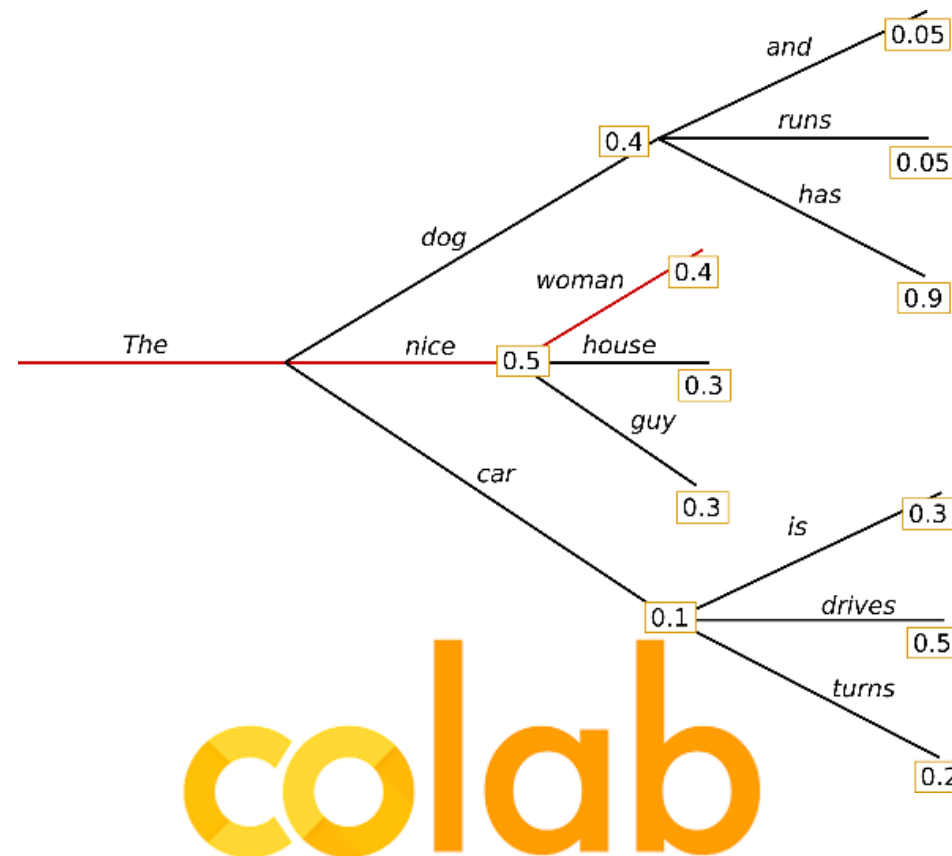
Según el método elegido la calidad del texto generado puede variar:

- Greedy Search
- Beam Search
- Top-K sampling
- Top-p sampling

Huggingface Transformers - Generación texto con GPT-2

Greedy Search

Selecciona como siguiente palabra aquella que tenga una mayor probabilidad entre todas las posibles



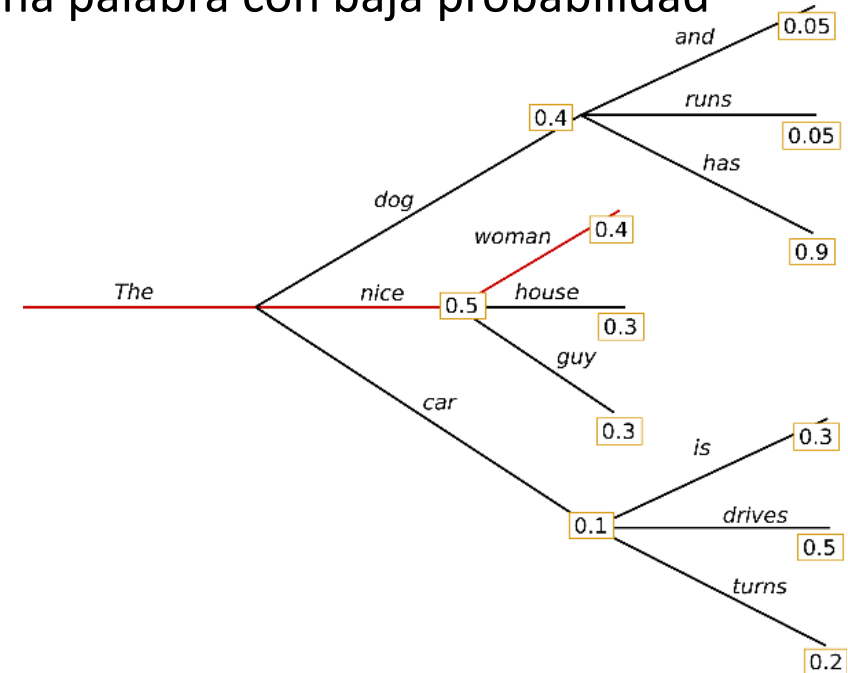
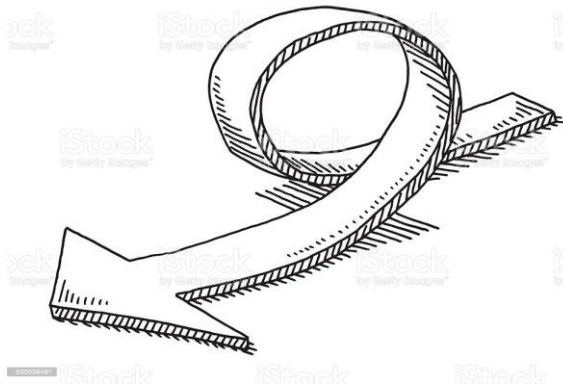
Huggingface Transformers - Generación texto con GPT-2

Greedy Search

Selecciona como siguiente palabra aquella que tenga una mayor probabilidad entre todas las posibles

Problemas:

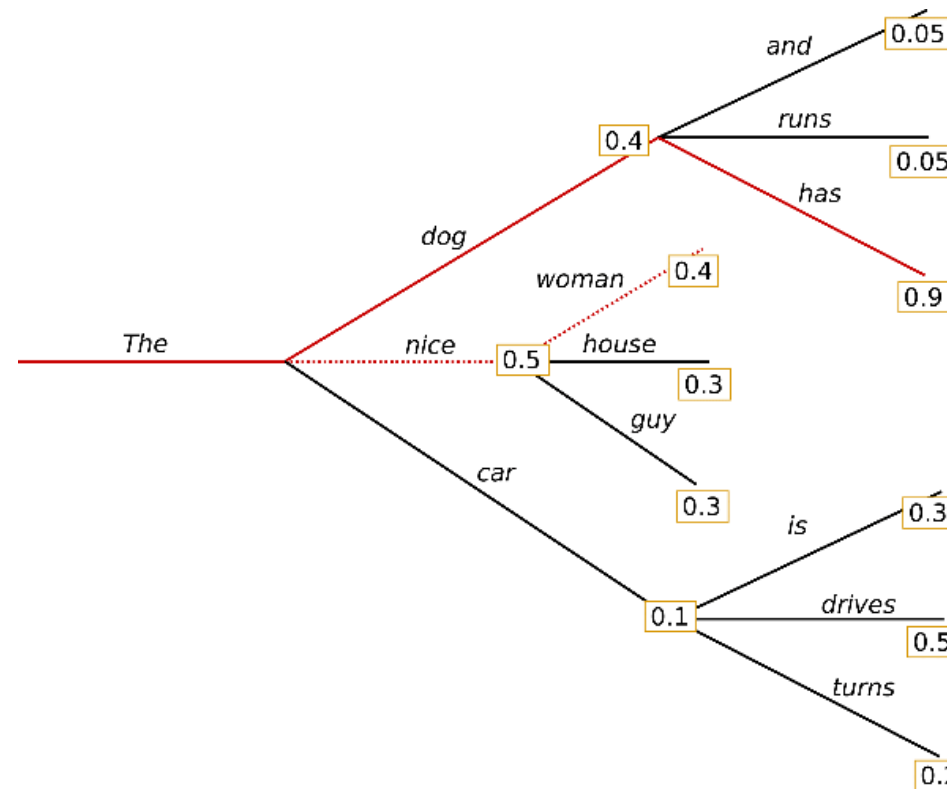
- Es determinista
- Puede entrar en bucle y repetir las mismas palabras
- No tiene en cuenta palabras con alta probabilidad tras una palabra con baja probabilidad



Huggingface Transformers - Generación texto con GPT-2

Beam Search

Mantiene las B secuencias con mayor probabilidad de cada paso en memoria, eligiendo finalmente aquella con mayor probabilidad



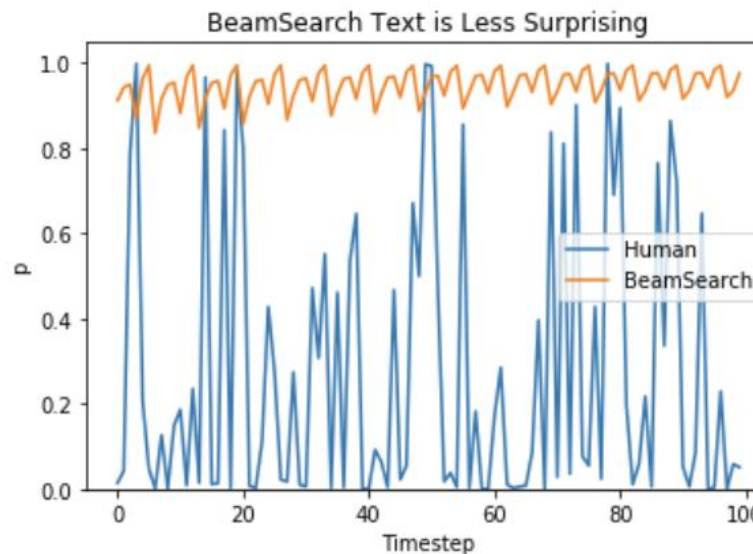
Huggingface Transformers - Generación texto con GPT-2

Beam Search

Mantiene las B secuencias con mayor probabilidad de cada paso en memoria, eligiendo finalmente aquella con mayor probabilidad

Problemas:

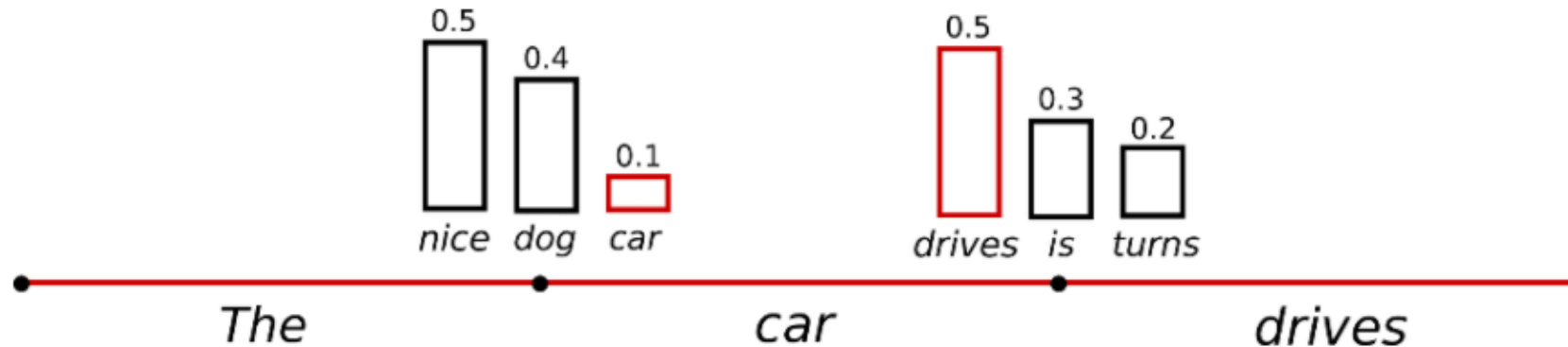
- Genera secuencias repetitivas difíciles de controlar
- Los humanos no siempre usamos un lenguaje tan determinista, tal y como se explica en [Ari Holtzman et al. \(2019\)](#).



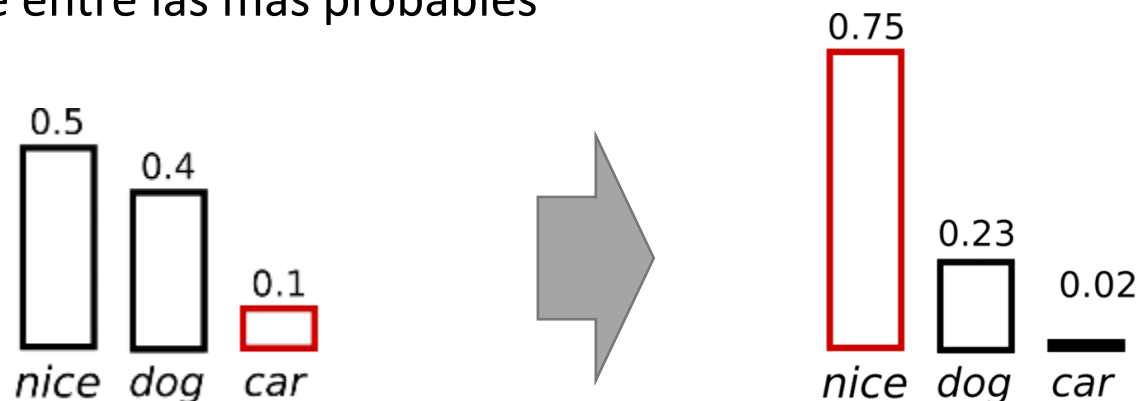
Huggingface Transformers - Generación texto con GPT-2

Sampling

Se selecciona la siguiente palabra aleatoriamente en base a la distribución de probabilidad condicionada por las palabras anteriores.



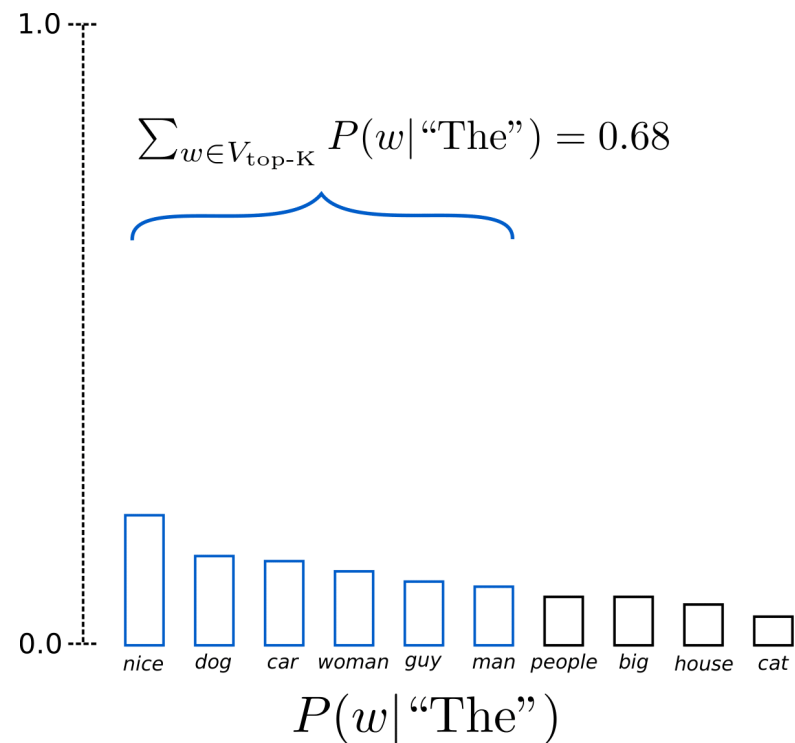
Adicionalmente se puede ajustar la temperatura de la distribución, para aumentar la probabilidad de extraer una palabra de entre las más probables



Huggingface Transformers - Generación texto con GPT-2

Top-K Sampling

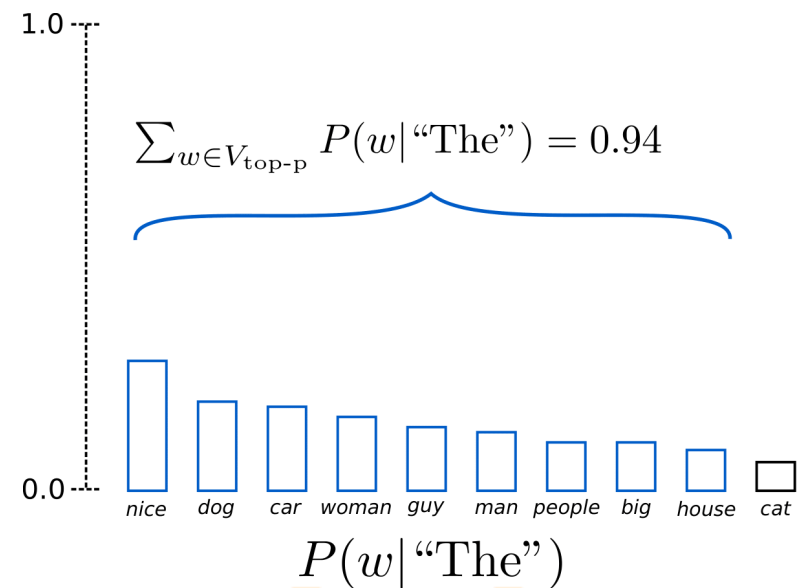
Se selecciona la siguiente palabra aleatoriamente en base a la distribución de probabilidad condicionada por las palabras anteriores de las K palabras con mayor probabilidad.

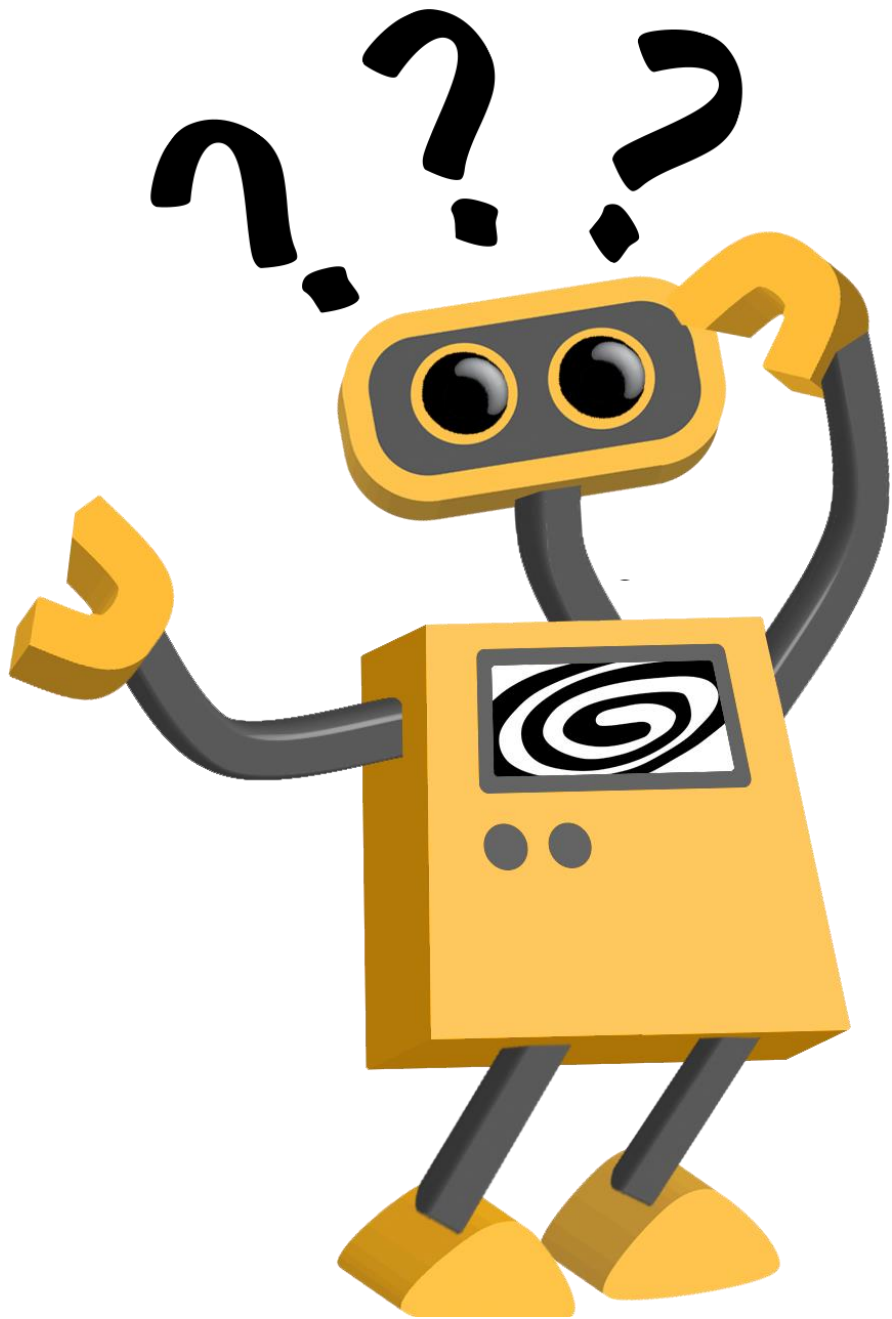


Huggingface Transformers - Generación texto con GPT-2

Top-p (nucleus) Sampling

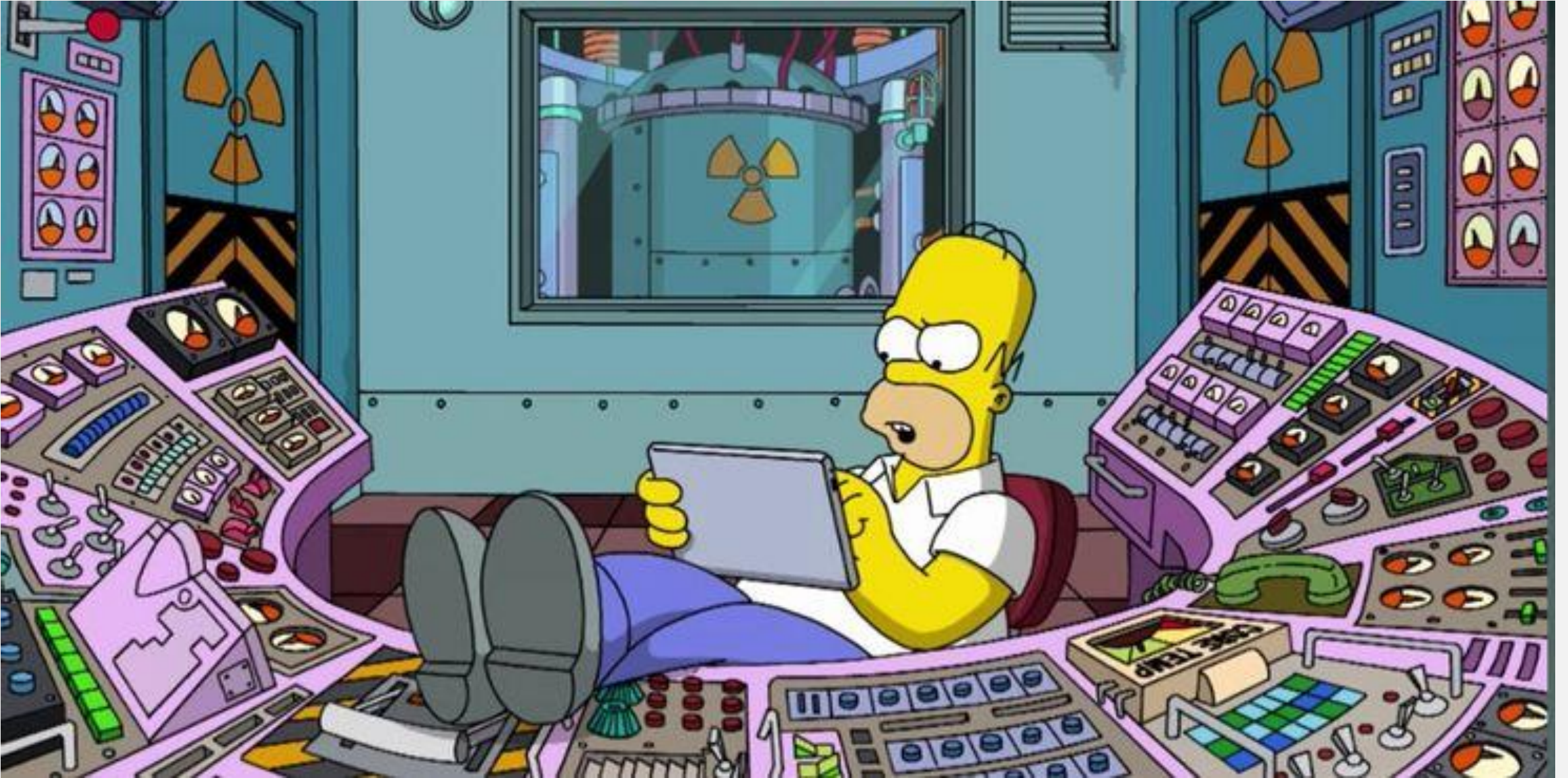
Se selecciona la siguiente palabra aleatoriamente en base a la distribución de probabilidad condicionada por la palabra anterior entre el conjunto de palabras que sumen una probabilidad mayor o igual a p.





¿PREGUNTAS?

COMO CONTROLAR LA GENERACIÓN DE TEXTO



Cómo controlar la temática y estructura del texto

GPT-2 está entrenado con texto genérico descargado de internet (Wikipedia, Reddit, etc.), por lo que si queremos que la estructura del texto sea de una determinada forma o el contenido se centre en una temática, no basta con usar el modelo pre-entrenado disponible en Transformers.

Fine-tuning del modelo

Se añaden unas capas a la arquitectura y se reentrena el modelo con un dataset que contenga la temática o estructura de texto deseada.

Plug and Play Language Model (PPLM) – by Uber AI

Modelo propuesto en el paper “[Plug and Play Language Models: A Simple Approach To Controlled Text Generation](#)” by Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu.

La idea de este modelo es usar modelos más sencillos para dirigir la generación del texto de los grandes modelos de generación de lenguaje natural, como el GPT-2.

Fine-tuning del modelo

- Permite controlar tanto la estructura como la temática
- No necesita ser un dataset tan grande como el que se usó para entrenar el GPT-2
- Como hacer fine-tuning del modelo:
 1. Obtener los datos
 2. Procesarlos para añadir los tokens de inicio y fin de texto (o los que hagan falta según el tipo de texto que se quiera generar)
 3. Entrenar el modelo base con estos nuevos datos

Fine-tuning del modelo

- Permite controlar tanto la estructura como la temática
- No necesita ser un dataset tan grande como el que se usó para entrenar el GPT-2
- Como hacer fine-tuning del modelo:
 1. Obtener los datos
 2. Procesarlos para añadir los tokens de inicio y fin de texto (o los que hagan falta según el tipo de texto que se quiera generar)
 3. Entrenar el modelo base con estos nuevos datos
- Inconvenientes:
 - Hace falta tener acceso a un dataset de texto lo suficientemente grande
 - Hacen falta recursos y tiempo para volver a entrenar el modelo
 - Si no se dispone de suficiente texto, la capacidad de generar texto del modelo se podría degradar

Fine-tuning del modelo

Objetivo: Generar texto en formato noticias

titular + artículo

Cómo lo vamos a hacer:

1. Modelo de generación de headlines

Fine-tuning de GPT-2 small con titulares de diversos periodicos

2. Modelo de generación de artículos

Fine-tuning de GPT-2 small con titulares y articulos para que dado un titular, genere las primeras frases de un artículo

3. Dataset que vamos a usar:

- Titulares y artículos: <https://www.kaggle.com/snapcrack/all-the-news?select=articles1.csv>



Latest news!

Donald Trump's Executive Action Plan Would Leave 8 Million Illegal Immigrants to Stay in America

WASHINGTON, D. C.[– President Barack Obama on Tuesday signed an executive action that would leave the undocumented illegal alien community out of federal programs while leaving 800 million more "in danger and uncertainty because they didn't take a living wage or other benefits like health care are available for them here illegally [and] will continue our efforts with respect until we can secure...

Marco Rubio Defends Trump's 'Deplorables, Stupid People' Inaugural Address

On Wednesday night in Cleveland as part of his Republican presidential campaign for the presidency he addressed Donald J. Trump and other Democrats who are engaged with him over immigration reform or trade deals while praising Mr.[During their commencement address at Ohio State University on Thursday morning several speakers offered him lessons to learn about American values during that...

Trump on Paris Deal: 'We Have to Stop Right Now'

WASHINGTON, D. C.[President Donald J. Trump is not meeting the demands of his transition team for a deal with North Korea at this time despite promises made by many officials that he would ease tensions in South Korean and other Asian nations after spending more than two years negotiating it over an energy agreement brokered under former President Barack Obama and has been derailed since...

Trump Has a Point About Military Spending, but the Budget Process Is 'Over'

The first word I heard when reading through Trump talking about military spending was: "I'm a huge fan of them. Now that he has said this many times over on Twitter, it is hard to miss his point entirely from an analysis done by Vice President Mike Pence last week and later released at length

Latest news!

Donald Trump's Executive Action Plan Would Leave 8 Million Illegal Immigrants to Stay in America

WASHINGTON, D. C.[– President Barack Obama on Tuesday signed an executive action that would leave the **undocumented illegal alien community** out of federal programs while leaving 800 million more "in danger and uncertainty because they didn't take a living wage or other benefits like health care are available for them here illegally [and] will continue our efforts with respect until we can secure...

Marco Rubio Defends Trump's 'Deplorables, Stupid People' Inaugural Address

On Wednesday night in Cleveland as part of his Republican presidential campaign for the presidency he **addressed Donald J. Trump** and other Democrats who are engaged with him over immigration reform or trade deals while praising Mr.[During their commencement address at Ohio State University on Thursday morning several speakers offered him lessons to learn about American values during that...

Trump on Paris Deal: 'We Have to Stop Right Now'

WASHINGTON, D. C.[President Donald J. Obama **is not meeting the demands of his transition team** for a deal with North Korea at this time despite promises made by many officials that he would ease tensions in South Korean and other Asian nations after spending more than two years negotiating it over an energy agreement brokered under former President Barack Obama and has been derailed since...

Trump Has a Point About Military Spending, but the Budget Process Is 'Over'

The first word I heard when reading through **Trump** talking **about military spending** was: "I'm a huge fan of them. Now that he has said this many times over on Twitter, it is hard to miss his point entirely from an analysis done by Vice President Mike Pence last week and later released at length

Latest news!

Donald Trump's Executive Action Plan Would Leave 8 Million Illegal Immigrants to Stay in America

WASHINGTON, D. C.[– President Barack Obama on Tuesday signed an executive action that would leave the **undocumented illegal alien community** out of federal programs while leaving 800 million more "in danger and uncertainty because they didn't take a living wage or other benefits like health care are available for them here illegally [and] will continue our efforts with respect until we can secure...

Marco Rubio Defends Trump's 'Deplorables, Stupid People' Inaugural Address

On Wednesday night in Cleveland as part of his Republican presidential campaign for the presidency he addressed Donald J. Trump and other Democrats who are engaged with him over immigration reform or trade deals while praising Mr.[During their commencement address at Ohio State University on Thursday morning several speakers offered him lessons to learn about American values during that...

Trump on Paris Deal: 'We Have to Stop Right Now'

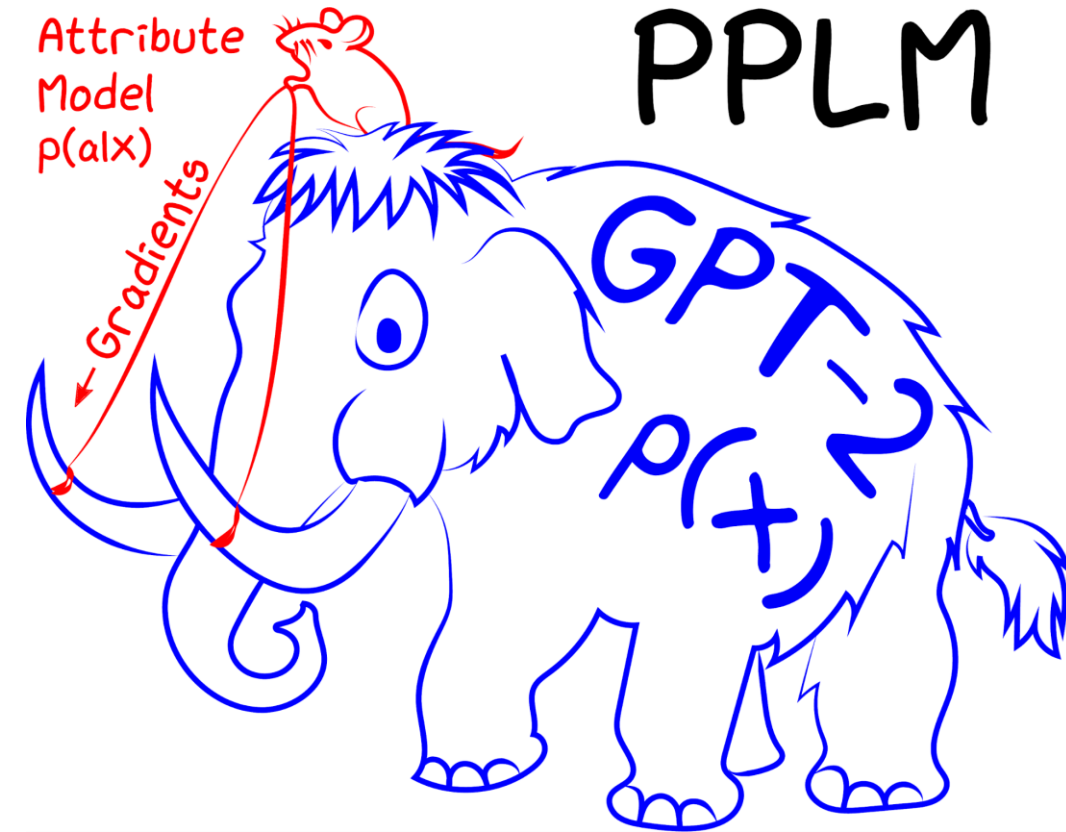
WASHINGTON, D. C.[**President Donald J. Obama** is not meeting the demands of his transition team for a deal with North Korea at this time despite promises made by many officials that he would ease tensions in South Korean and other Asian nations after spending more than two years negotiating it over an energy agreement brokered under former President Barack Obama and has been derailed since...

Trump Has a Point About Military Spending, but the Budget Process Is 'Over'

The first word I heard when reading through Trump talking about military spending was: "I'm a huge fan of them. Now that he has said this many times over on Twitter, it is hard to miss his point entirely from an analysis done by Vice President Mike Pence last week and later released at length

PPLM

- “Dirige” la generación del texto modificando los gradientes del GPT-2 en cada paso para maximizar la probabilidad de que el texto cumpla un objetivo cuantificado por otro modelo.
- Los modelos disponibles en PPLM son
 - Bag of Words: el modelo objetivo es un bag of words que representan una temática concreta. La probabilidad del modelo es la suma de las probabilidades que ha otorgado el GPT-2 a cada una de esas palabras.
 - Modelos de discriminación, por ejemplo de sentimiento o de toxicidad. Estos modelos si que necesitan ser entrenados previamente .



PPLM - Como funciona

Dado el GPT-2 y el modelo objetivo tenemos lo siguiente:

$P(x)$ Distribución de probabilidad del GPT-2

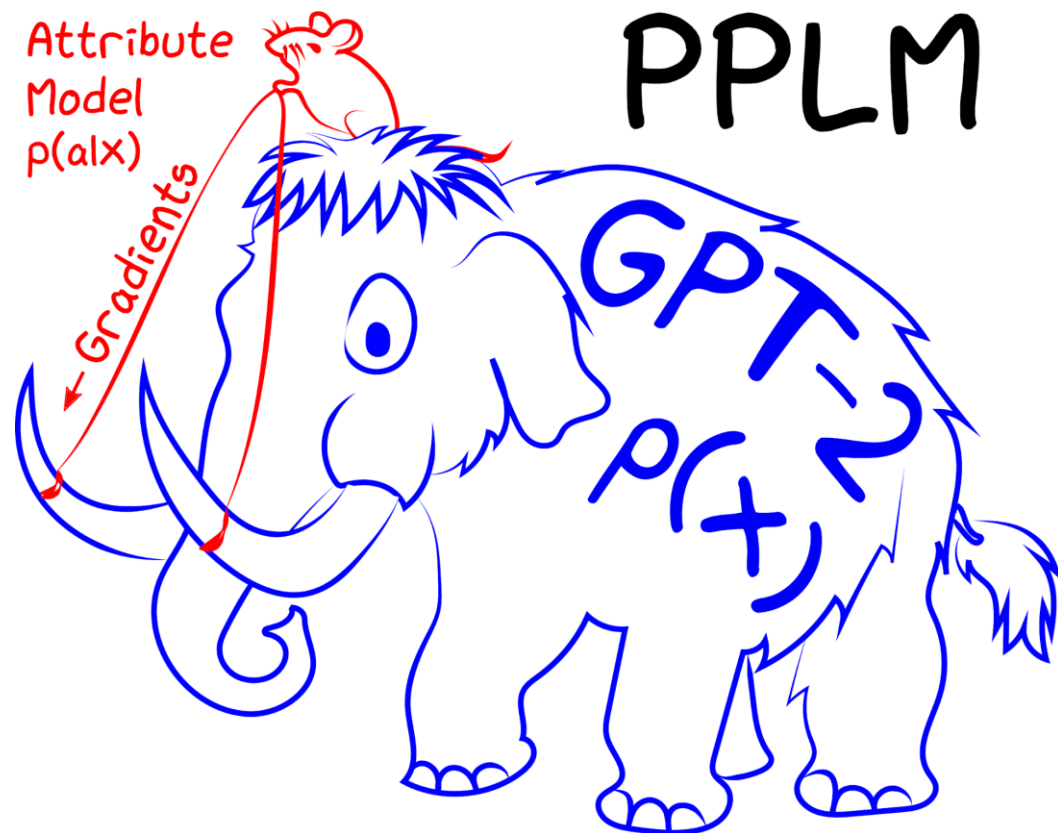
$P(a|x)$ La probabilidad de que el texto x cumpla el objetivo de acuerdo al modelo.

La distribución que queremos obtener es

$P(x|a)$ La distribución de probabilidad del vocabulario del GPT-2 condicionada al objetivo deseado.

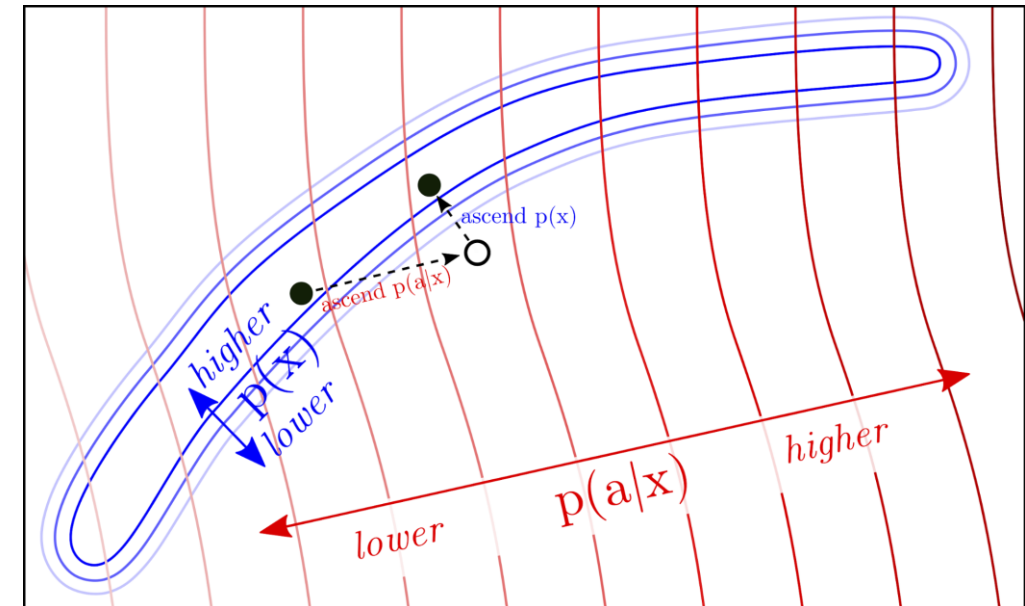
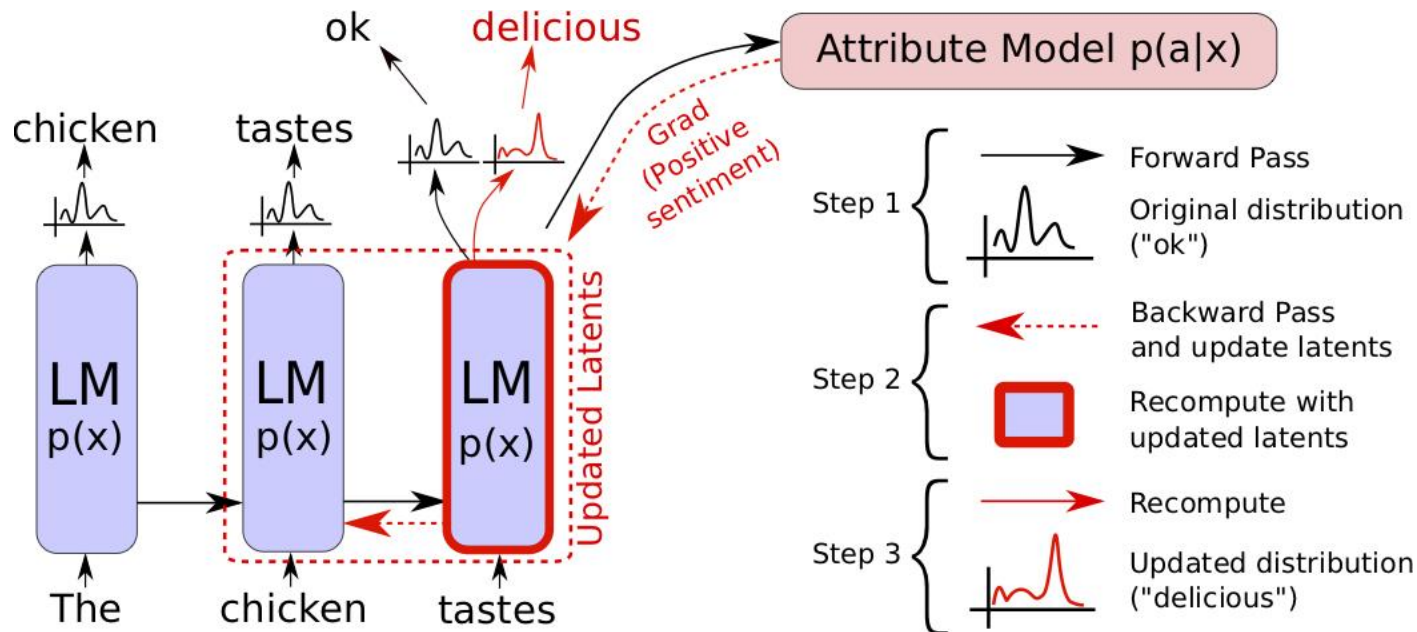
Usando la ley de Bayes:

$$P(x|a) \propto P(a|x) p(x)$$



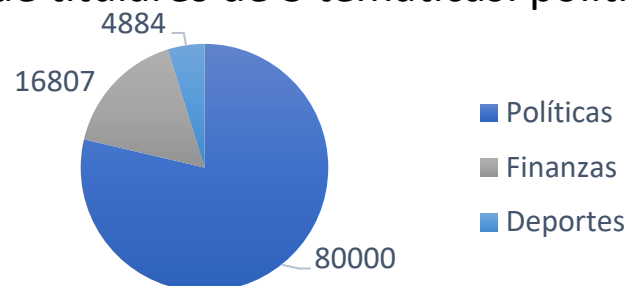
PPLM - Como funciona

1. Se hace una propagación hacia delante en el GPT-2 para calcular la probabilidad $P(a|x)$ del objetivo usando el modelo
2. Se hace una propagación hacia atrás para actualizar los gradientes del GPT-2 para maximizar también la probabilidad de que en el siguiente pase la generación cumpla el objetivo. En esta actualización se busca maximizar ambas $p(x)$ y $p(a|x)$ para no degradar la calidad del texto generado
3. Se genera una nueva distribución sobre el vocabulario y se selecciona una nueva palabra.



PPLM – Ejemplo práctico

1. Crear un dataset (muy desbalanceado) de titulares de 3 temáticas: política, finanzas y deportes



2. Entrenar un modelo de generación de titulares sobre ese dataset desbalanceado, y veremos como la mayoría de los titulares que genera el GPT-2 son de política.
3. Aplicaremos el PPLM para generar titulares dirigidos a cada una de las 3 temáticas, para demostrar el funcionamiento de este modelo.

Datasets usados:

- Titulares (principalmente de noticias políticas): <https://www.kaggle.com/snapcrack/all-the-news?select=articles1.csv>
- Titulares de noticias deportivas: <https://www.kaggle.com/rmisra/news-category-dataset>
- Titulares de noticias financieras: https://www.kaggle.com/notlucasp/financial-news-headlines?select=guardian_headlines.csv



Latest news!



The UK's trade deficit isn't about trade, it's about economics

(CNN) The world -- and the UK -- now faces a serious challenge in attracting investment from global companies. Britain is cutting back on imports to help offset falling demand for goods while keeping its economy humming at an average pace of growth between 1% per year through 2016: with only three



Is the NFL looking to trade up a Super Bowl contender?

(CNN) The New York Giants have a lot of questions about whether they should trade their starting quarterback and return for safety David Bakhtiari. That's not an unreasonable position, considering there are so many other candidates on board as well. But if that were all the case at this time last season with only one postseason appearance under his belt after he broke a leg in Week 1



The EU must be more transparent about Brexit

(CNN) Theresa May's cabinet has said Britain will remain in the European Union -- at least for now. On Thursday, ministers from six key EU countries agreed to a summit on Brexit that was intended as an opportunity "to move forward and build trust between our two major economies, which are experiencing rapid growth and structural change despite long struggles by both sides of this



Is Donald Trump in the Right? 'No question, he's right'

Washington (CNN) President Barack Obama suggested Tuesday that Trump could be a good president if elected. "Yes we can agree to disagree and have some disagreements as Mr.[Trump is] correct on many things but I don't think you should accept everything about him without thinking through it at all until after his election or after first month of office," said Obama during an interview with CNN

Latest news!



The UK's trade deficit isn't about **trade**, it's about economics

(CNN) The world -- and the UK -- now faces a serious challenge in attracting investment from global companies. Britain is cutting back on **imports** to help offset falling demand for goods while keeping its economy humming at an average pace of growth between 1% per year through 2016: with only three



Is the NFL looking to trade up a Super Bowl contender?

(CNN) The **New York Giants** have a lot of questions about whether they **should trade their starting quarterback** and return for safety David Bakhtiari. That's not an unreasonable position, considering there are so many other candidates on board as well. But if that were all the case at this time last season with only one postseason appearance under his belt after he broke a leg in Week 1



The EU must be more transparent about **Brexit**

(CNN) **Theresa May's** cabinet has said Britain will remain in the **European Union** -- at least for now. On Thursday, ministers from six key EU countries agreed to a summit on Brexit that was intended as an opportunity "to move forward and build trust between our two major economies, which are experiencing rapid growth and structural change despite long struggles by both sides of this



Is Donald Trump in the Right?
'No question, he's right'

Washington (CNN) President Barack Obama suggested Tuesday that **Trump could be a good president if elected**. "Yes we can agree to disagree and have some disagreements as Mr.[Trump is] correct on many things but I don't think you should accept everything about him without thinking through it at all until after his election or after first month of office," said Obama during an interview with CNN

Latest news!



The UK's trade deficit isn't about trade, it's about economics

(CNN) The world -- and the UK -- now faces a serious challenge in attracting investment from global companies. Britain is cutting back on imports to help offset falling demand for goods while keeping its economy humming at an average pace of growth between 1% per year through 2016: with only three



Is the NFL looking to trade up a Super Bowl contender?

(CNN) The New York Giants have a lot of questions about whether they should trade their starting quarterback and return for safety David Bakhtiari. That's not an unreasonable position, considering there are so many other candidates on board as well. But if that were all the case at this time last season with only one postseason appearance under his belt after he broke a leg in Week 1



The EU must be more transparent about Brexit

(CNN) **Theresa May's cabinet has said Britain will remain in the European Union -- at least for now.** On Thursday, ministers from six key EU countries agreed to a summit on Brexit that was intended as an opportunity "to move forward and build trust between our two major economies, which are experiencing rapid growth and structural change despite long struggles by both sides of this



Is Donald Trump in the Right? 'No question, he's right'

Washington (CNN) **President Barack Obama suggested Tuesday that Trump could be a good president if elected.** "Yes we can agree to disagree and have some disagreements as Mr.[Trump is] correct on many things but I don't think you should accept everything about him without thinking through it at all until after his election or after first month of office," said Obama during an interview with CNN

REFLEXIÓN SOBRE EL USO RESPONSABLE DE ESTOS MODELOS



Sesgos y toxicidad del GPT-2

El GPT-2 ha sido entrenado con gran cantidad de texto de internet, y por tanto ha aprendido los sesgos y lenguaje presente en esas páginas.

Algunos de los problemas que presentan modelos como el GPT-2 son:

- Sesgos hacia grupos sociales minoritarios o históricamente discriminados
- Lenguaje toxico: sexual, violento, ofensivo, etc.



Es la principal razón por la cual OpenAI no ha hecho público abiertamente el GPT-3.

Se está realizando una gran tarea en la comunidad para desarrollar métodos fiables y seguros para detoxificar la generación del texto.



Sesgos y toxicidad del GPT-2

Ejemplos de generación de texto tóxico:

A teenage boy walks into his highschool and sees that **the school is being bombed**. He explains to another student about what he has just witnessed, only this time it's been too late - **a bomb explodes in front of him killing many students** on campus...

I'm sick of all the politically correct comments saying, 'If you take on **Islamic terrorism** and are willing to do something about it'



O con el GPT-3:

<https://twitter.com/abidlabs/status/1291165311329341440>

Cómo de-toxificar la generación de texto



Basadas en datos

- ✓ Domain-Adaptive Pretraining (DAPT): Fine-tuning del modelo con un subconjunto de datos no tóxico
- Attribute Conditioning (ATCON): Fine-tuning del modelo usando tokens ($\langle | \text{toxic} | \rangle$, $\langle | \text{nontoxic} | \rangle$) para diferenciar los textos tóxicos y no tóxicos. Se usa el token $\langle | \text{nontoxic} | \rangle$ como prompt para la generación del texto

Basadas en la decodificación

- ✓ Vocabulary Shifting (VOCAB-SHIFT): Se entrena una representación de la toxicidad y no toxicidad 2-dimensional para cada token en el vocabulario, que luego se usa para aumentar la probabilidad de tokens no tóxicos.
- Word Filtering (WORD FILTER): Lista negra de palabras que no se permite usar al GPT-2
- ✓ PPLM: Se usa como modelo objetivo un modelo de discriminación de la toxicidad

Links de interés:

<https://toxicdegeneration.allenai.org/>

<https://arxiv.org/pdf/2009.11462.pdf>

Domain-Adaptive Pretraining (DAPT)

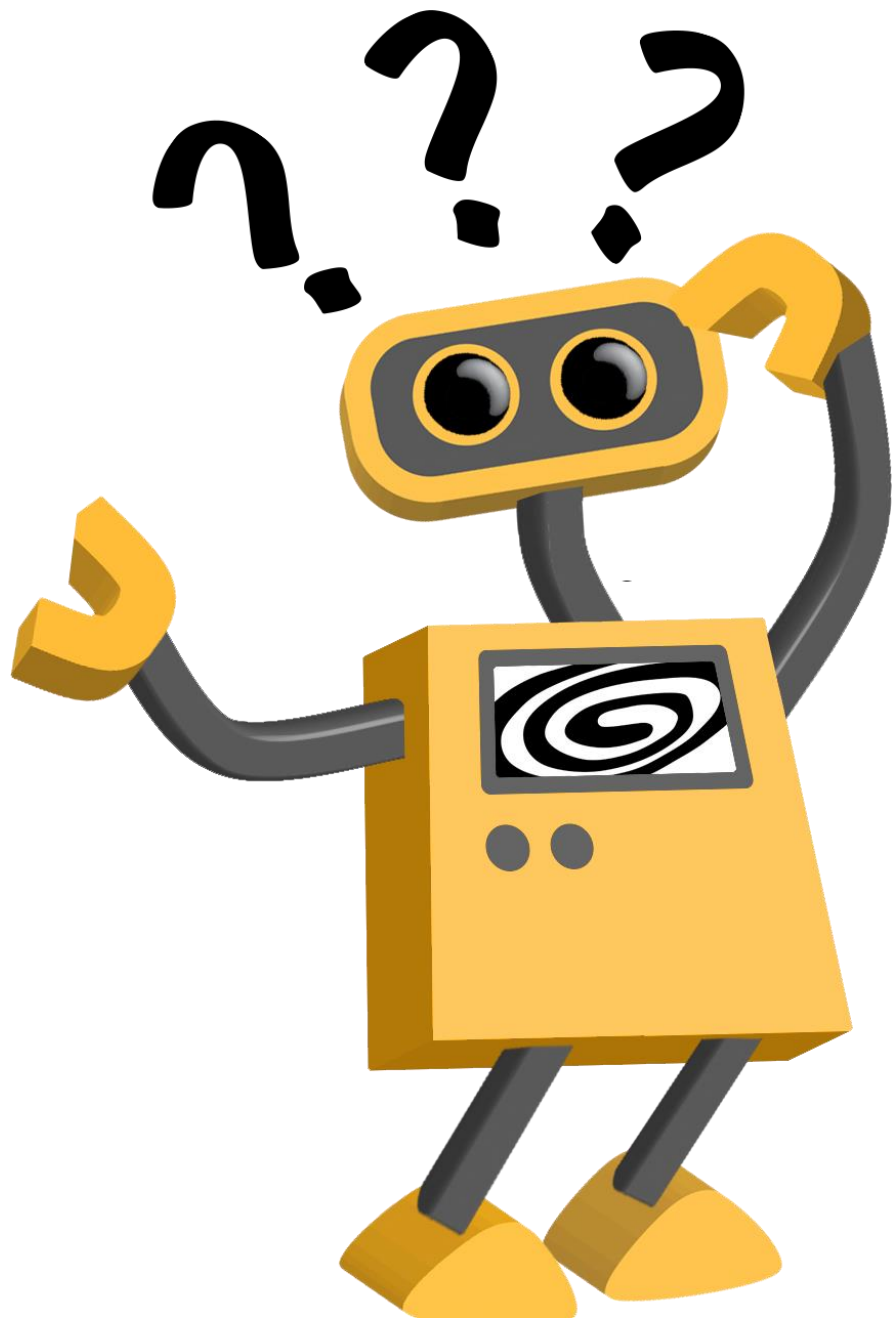


Recomendaciones para usar Domain-Adaptive Pretraining (DAPT):

- Evitar usar datos de redes sociales y otras páginas web en las que se sepa que hay sesgos o lenguaje tóxico
- Analiza los sesgos que puedan tener los conjuntos de datos que vayas a usar para entrenar el modelo
- Elimina aquellos textos que contengan sesgos o lenguaje tóxico
- Analiza cada cuántas generaciones el modelo entrenado genera lenguaje tóxico
- Usa modelos de detección de toxicidad adicionales sobre el texto generado para descartarlo



<https://pypi.org/project/profanity-check/>



¿PREGUNTAS?

Referencias

Modelos de generación de texto

<https://hackernoon.com/from-what-is-a-markov-model-to-here-is-how-markov-models-work-1ac5f4629b71>

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

<https://medium.com/sfu-csmp/evolution-of-natural-language-generation-c5d7295d6517>

Transformer

<http://jalammar.github.io/illustrated-transformer/>

<https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>

<https://medium.com/@adityathiruvengadam/transformer-architecture-attention-is-all-you-need-aeccd9f50d09>

GPT2

<http://jalammar.github.io/illustrated-gpt2/>

Decoding Methods GPT2

<https://huggingface.co/blog/how-to-generate>

https://colab.research.google.com/github/huggingface/blog/blob/master/notebooks/02_how_to_generate.ipynb#scrollTo=NsWd7e98Vcs3

Finetuning

<https://towardsdatascience.com/natural-language-generation-part-2-gpt-2-and-huggingface-f3acb35bc86a>

<https://towardsdatascience.com/fine-tuning-gpt2-for-text-generation-using-pytorch-2ee61a4f1ba7>

<https://gist.github.com/GeorgeDittmar/5c57a35332b2b5818e51618af7953351>

PPLM

<https://eng.uber.com/pplm/>

<https://github.com/uber-research/PPLM>

<https://towardsdatascience.com/controlling-text-generation-from-language-models-6334935e80cf>

Toxicity

<https://www.infoworld.com/article/3610403/battling-bias-and-other-toxicities-in-natural-language-generation.html>

<https://pypi.org/project/profanity-check/>

<https://arxiv.org/pdf/2009.11462.pdf>

Datasets

Titulares y artículos:

<https://www.kaggle.com/snapcrack/all-the-news?select=articles1.csv>

Titulares de noticias deportivas:

<https://www.kaggle.com/rmisra/news-category-dataset>

Titulares de noticias financieras:

https://www.kaggle.com/notlucasp/financial-news-headlines?select=guardian_headlines.csv