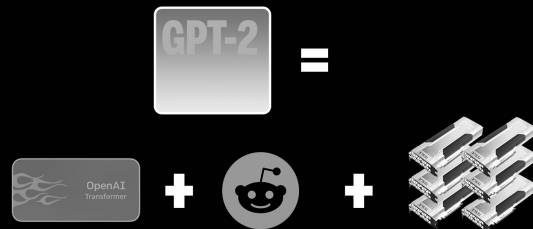Essam Babu

DeepGhostwriter: Natural Language Generation, Transformers, and Hip-Hop

"Men lie, women lie, numbers don't" - Jay Z

GPT-2 =

OpenAI Transformer + reddit +

Introduction: Transformers

- Improves on RNNs + CNNs

- Stacks of **Encoders** and **Decoders** linked together

  - **Encoder**: process its input (word vector) to generate **encodings**, containing **information about which parts of the inputs are relevant to each other**. It passes its set of encodings to the next encoder layer as inputs

  - **Decoder**: Does the opposite, taking all the encodings and processes them, using their **incorporated contextual information** to **generate an output sequence**

- **Attention Mechanism**: Allows us to **focus on relevant past items** in the sequence

  Attention Is All You Need (Vaswani et al. 2017)
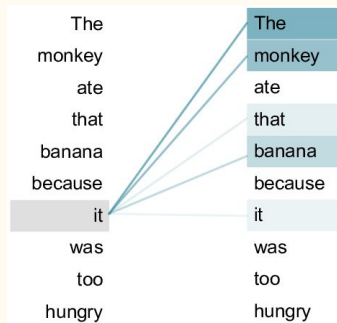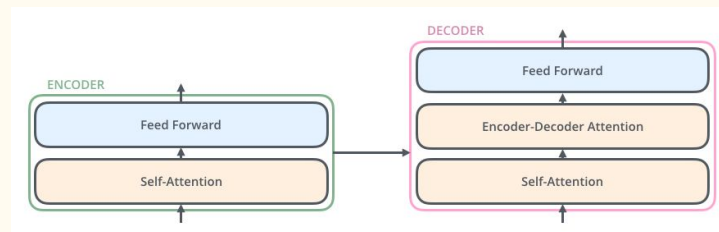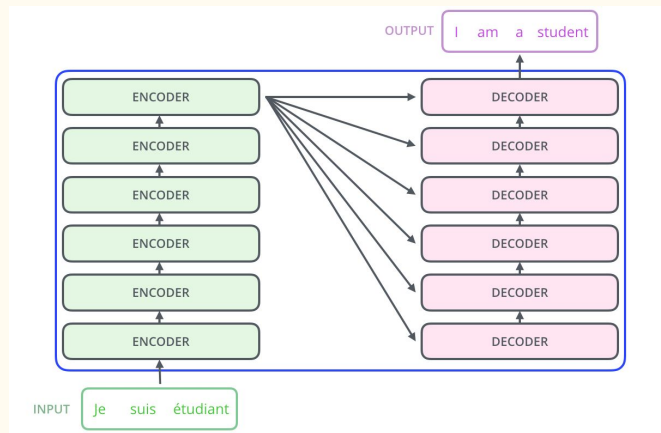
GPT2

GPT-2 (OpenAI, 2019) is a transformers-based language model trained on a very large corpus of English data in a **self-supervised** way.

- Raw texts only, with no humans labelling

- More precisely, it was trained to **guess the next word** in sentences using **masked self-attention**

- You can then use **sampling techniques** to transform the Decoder's embedding sequence into actual text outputs.
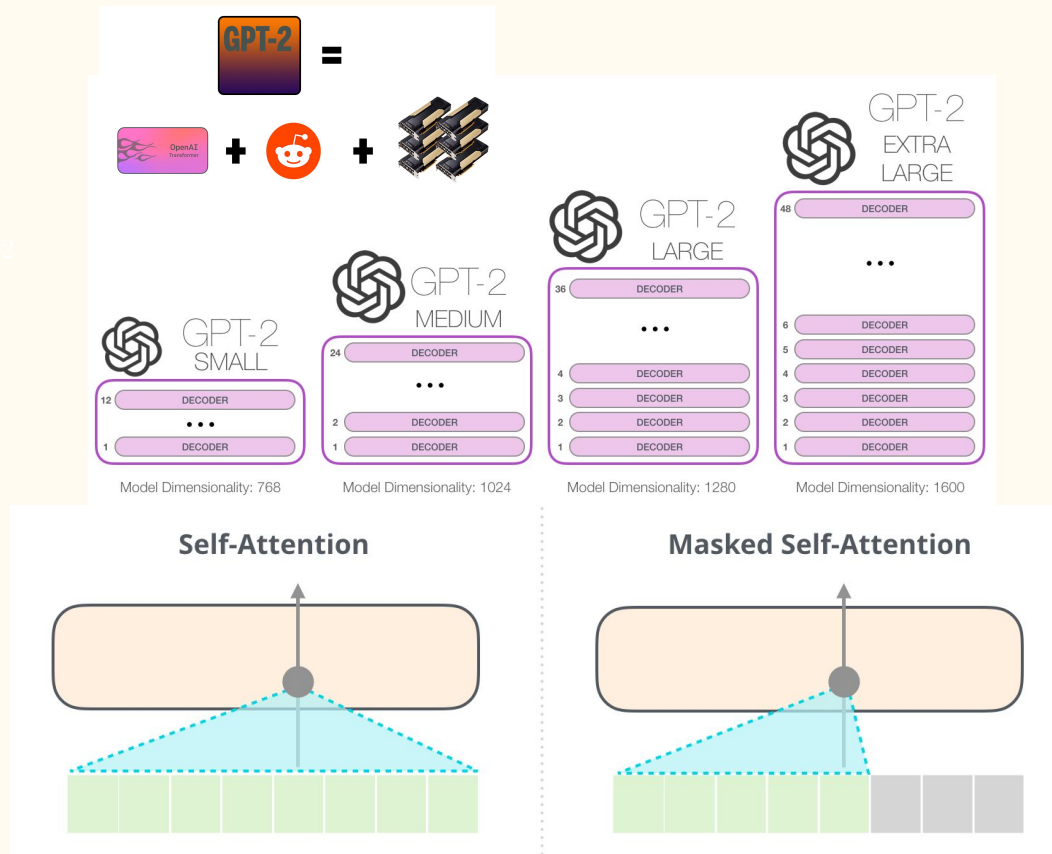
Transfer Learning

"Don't reinvent the wheel."

We can take a pre-trained Transformer model and further train it on a refined set of data to perform a specific task such as:

- Summarization
- Question-and-Answer Dialogue
- Text generation

Methodology

- Pre-trained GPT-2 model ("Small", 117 M)
  - PyTorch + HuggingFace Transformers Library

- Data: Scraped 5700+ song lyrics from 143 artists (~40 songs each) from Genius.com using their API.
  - **Two different tokenized formats:**
    - <|startoftext|> LYRICS OF ONE SONG <|endoftext|>

    - <|startoftext|>
      <<Artist - Song Title>>
      [Verse 1]
      LYRICS
      [Chorus]

      ...
      <|endoftext|>

Generating Text by Sampling (top-K sampling)

$$K = 6$$

Top-K Sampling

- the K most likely next words are filtered and the probability mass is redistributed among only those K next words.

$\sum_{w \in V_{\text{top-K}}} P(w|\text{"The"}) = 0.68$

$\sum_{w \in V_{\text{top-K}}} P(w|\text{"The"}, \text{"car"}) = 0.99$

nice dog car woman guy man people big house cat

drives is turns stops down a not the small told

$P(w|\text{"The"})$

$P(w|\text{"The"}, \text{"car"})$
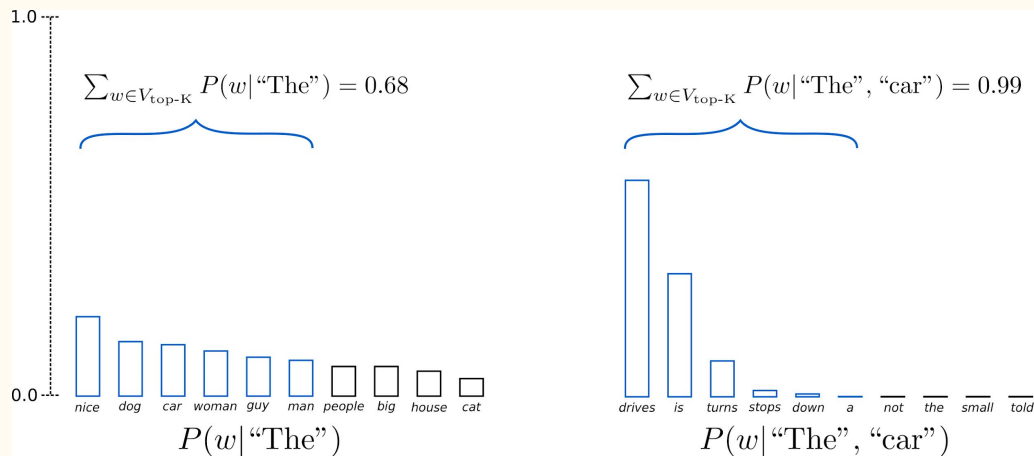
Generating Text by Sampling (top-P sampling)

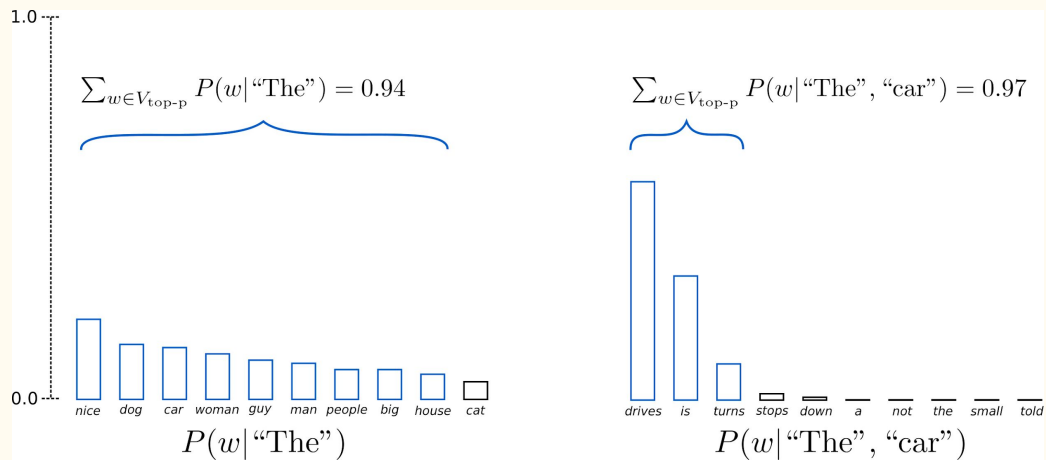**Top-p (nucleus) sampling**

- in Top-p sampling chooses from the smallest possible set of words whose cumulative probability exceeds the probability *p*.

- The probability mass is then redistributed among this set of words. This way, the size of the set of words (a.k.a the number of words in the set) can **dynamically increase and decrease according to the next word's probability distribution.**

P=0.92

$$\sum_{w \in V_{\text{top-p}}} P(w|\text{"The"}) = 0.94$$

$$\sum_{w \in V_{\text{top-p}}} P(w|\text{"The"}, \text{"car"}) = 0.97$$

nice dog car woman guy man people big house cat

$P(w|\text{"The"})$

drives is turns stops down a not the small told

$P(w|\text{"The"}, \text{"car"})$

Results

The Heart and the Tongue - Chance the Rapper

Original Lyrics:

My heart and tongue are fighting, my mind is undecided
It's not like Trump and Biden, it's more like something private
Like when your cousins fighting, one of 'em get excited
You can't just jump the gun and pick a side and jump inside it

Remember, Pontius Pilate hung a God and justified it
The truth is slow 'cause someone's always in a rush to hide it
The lies is golden 'cause the Devil got a touch of Midas
The soul is brighter 'cause the Holy Ghost is ultraviolet
The tongue is violent when the spirit starts to hunger-striking
The thunder lightning makes the heart go under bunkers hiding
The lungs is silent when it hears the footsteps of the giants
And then, the fingertips, they twist up like Poseidon's trident
Remember safety if you ever running from the tight end
Remember Mookie if they ever come to cut the hydrant
They cut the heart out with the tongue, they want it undivided
You better not start, you make a fuss you end up uninvited

GPT2-Ghostwriter Output

p= 0.92, k=50

<|startoftext|>
My heart and tongue are fighting, my mind is undecided
It's not like Trump and Biden, it's more like something private
Like when your cousins fighting, one of 'em get excited
You can't just jump the gun and pick a side and jump inside it

It's getting harder and harder to see what's going on
There's a thin line between love and hate, it's not like you and me
There's a thin line between the two, there's a thin line between us
The world is getting too small for us to feel no pain
And it's been a long, long time, and I guess it's all about you
We're not alone if you're not feeling yourself
We're not just anybody
But we are a reflection of what we have to feel
And I am here for you and I hope that you feel this
'Cause if you don't, it's gon' be painful and you gon' regret it
When you get high, then you gotta learn that there's no one above you
When you wake up in the morning, get dressed and go to bed

Results

The Heart and the Tongue - Chance the Rapper

Original Lyrics:

[Verse 1]
My heart and tongue are fighting, my mind is undecided
It's not like Trump and Biden, it's more like something private
Like when your cousins fighting, one of 'em get excited
You can't just jump the gun and pick a side and jump inside it

Remember, Pontius Pilate hung a God and justified it
The truth is slow 'cause someone's always in a rush to hide it
The lies is golden 'cause the Devil got a touch of Midas
The soul is brighter 'cause the Holy Ghost is ultraviolet
The tongue is violent when the spirit starts to hunger-striking
The thunder lightning makes the heart go under bunkers hiding
The lungs is silent when it hears the footsteps of the giants
And then, the fingertips, they twist up like Poseidon's trident
Remember safety if you ever running from the tight end
Remember Mookie if they ever come to cut the hydrant
They cut the heart out with the tongue, they want it undivided
You better not start, you make a fuss you end up uninvited

GPT2-Ghostwriter Output

<|startoftext|>
[Verse 1]
My heart and tongue are fighting, my mind is undecided
It's not like Trump and Biden, it's more like something private
Like when your cousins fighting, one of 'em get excited
You can't just jump the gun and pick a side and jump inside it

Like we in the back of the '80s, they tell me it's time to "Go for It"
We all know this is real life, but I can't pretend
Like this isn't what it's 'bout, I can't pretend
Or pretend at all, we been living in this house for too long
And every time we do it, we say it, we celebrate it
We done put the city on our backs, we in it for self
I swear this ain't what it's 'bout
My heart is beating, my tongue is twitching
My heart is beating, my tongue is twitching

[Hook]
We be living this life, living this life
Living this life, living this life
Living this life, living this life
Staring at the stars in the heavens
Staring at the stars in the heavens

Evaluation: Rhyme Density and Unique Word Count

Evaluation: Rhyme Density and **Unique Word Count** of Original Song vs. Generated Lyrics

Assonance Rhyme: rhymes of vowels despite different consonant sounds.

My heart and tongue are fighting, my mind is undecided
It's not like Trump and Biden it's more like something private

Rhyme Density (Malmi et al. 2016):

- Run the lyrics transcript through g2p (grapheme-to-phoneme) to get **a phonetic version**. **Remove non-vowels.**

- For each word:
  - Find the **longest matching vowel sequence** near the word. (This is a **multisyllabic assonance rhyme**.)

- **Compute the density by averaging the lengths of the longest matching sequence over all words.**

of Unique Words Used Within Artist's First 35,000 lyrics
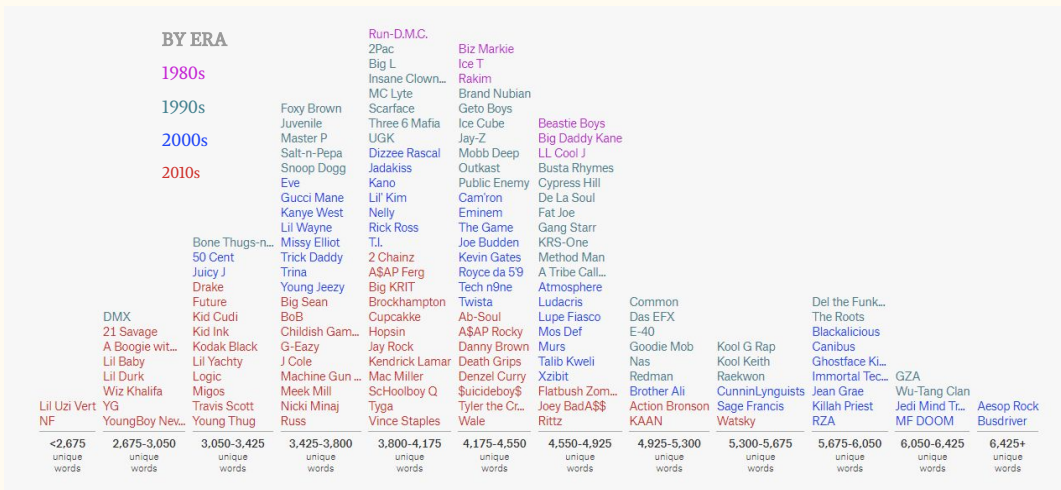
Matt Daniel (2019): Rappers sorted by size of their vocabulary.
https://pudding.cool/projects/vocabulary/index.html

Discussion

The model has somewhat learned the structure of hip-hop lyrics

Limitations:

- Low Rhyme Density
- Often ends up repeating sentences
- Very little semantic coherence

Extensions:

- Measure semantic similarity
- GPT-3 (175 B) and beyond?
- Problems learning from large amounts of raw text?

Thanks for listening!