



Conversational



LLMs

Uriel Cabañas
Pau Colomer
Martí Esquius



Table of contents

01

Motivation

Introduction and
motivation

02

State of the Art

LMs and LLMs architecture
overview

03

Methodology

Review of our experiments
and observations

04

Results

Metric exploration and
fine-tune evaluation

05

Conclusions

Review and final reflexion
over the project

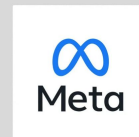
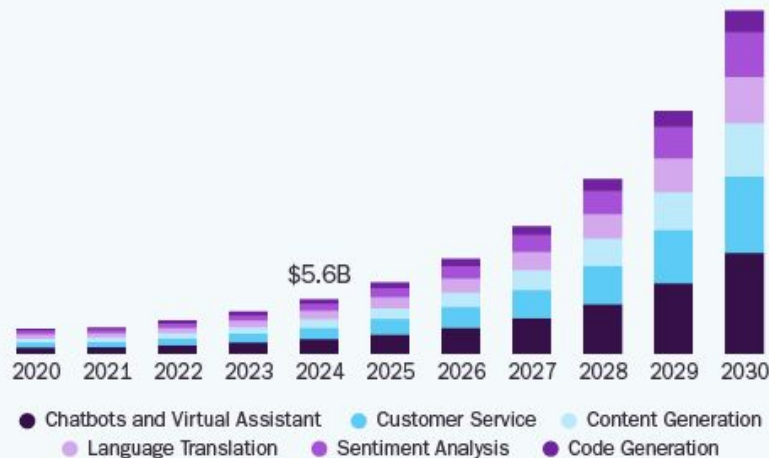


01 ★ Motivation ★

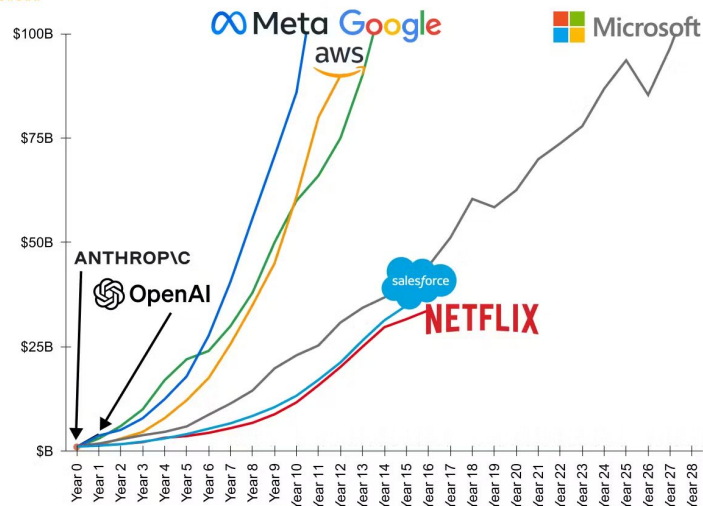
LLMs Overview

Large Language Models Market

Size, by Application, 2020 - 2030 (USD Billion)



Years from \$1B to \$100B revenue





+ 1M users

In just 5 days after launching

+ 400M WAU

Current Weekly Active Users (WAU)

\$300 Billion

OpenAI current valuation



+ 1.800.000.000.000

ChatGPT-4 parameters

Study objectives



Understand LLMs
architecture



Fine-Tuning

Explore Fine-Tuning techniques to
convert a modern LLM into a
poeth



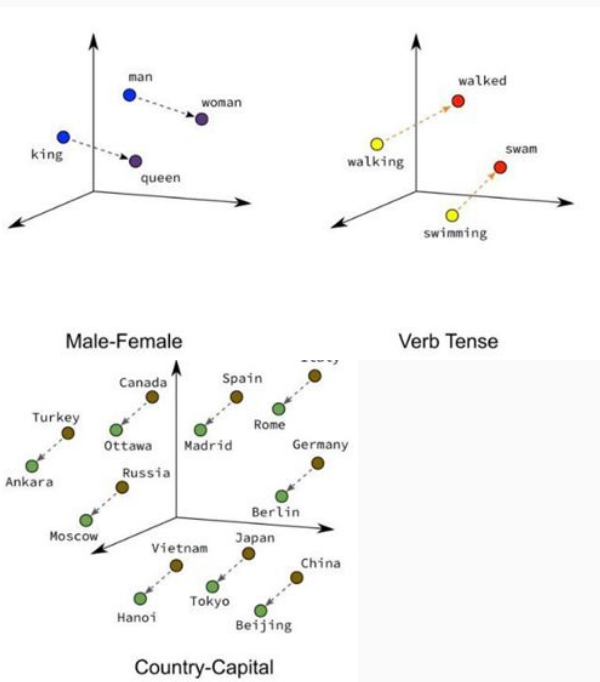
Explore efficient
strategies



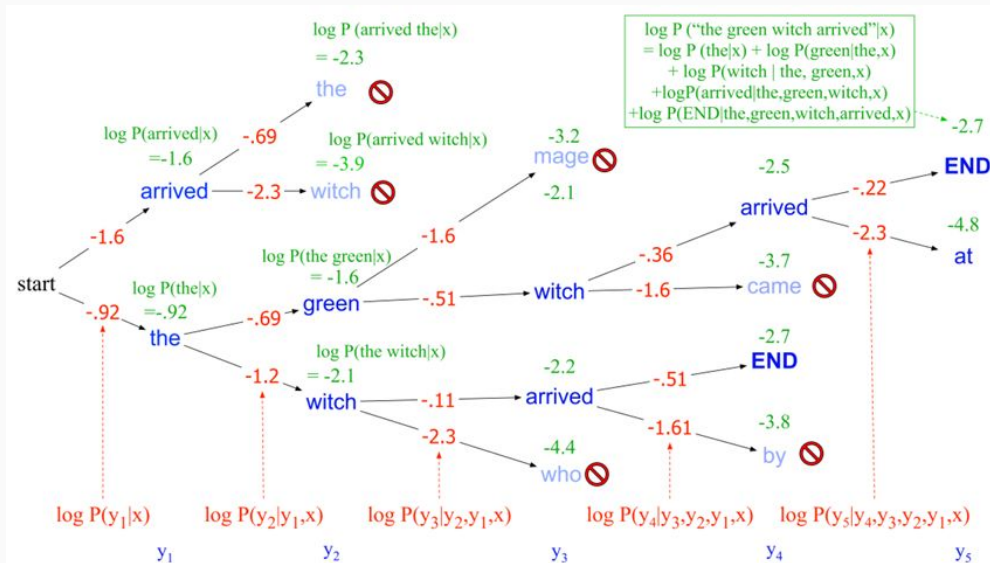
02 ✨ State of the Art

Language Models overview

Word Embeddings



Beam Search

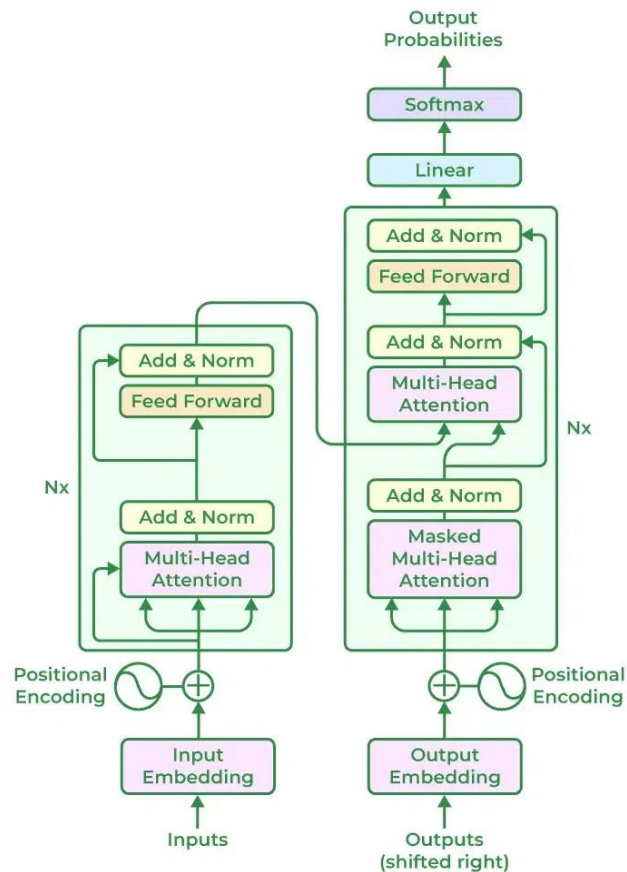


ChatGPT's Architecture

Transformer Architecture →

Other strategies

- Tokenization + Positional Encoders
- Transformer Layers
- Reinforcement Learning
- Pre-training + Fine-tuning





03 ✨ Methodology

Pre-trained Models



deepseek-ai/deepseek-coder-1.3b-base

- 1.3 Billion Parameters



ChatGPT

openai-community/gpt2

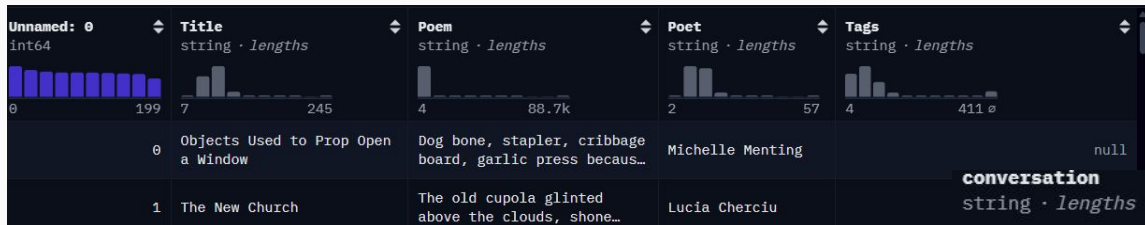
- 124 Million Parameters

- Decoder-only transformers – causal language modeling
- Byte-Pair-Encoding tokenizers

DataSet

For Fine-tuning

suayptalha/Poetry-Foundation-Poems



```
return {  
    "prompt": prompt,  
    "completion": completion  
}
```

isaacrehg/poetry-instructions

```
['prompt', 'completion']  
{ 'prompt': 'Can you write me a poem about faults and love?',  
  'completion': "Sure, here's a poem about faults
```

Pre-trained Models



conversation
string · lengths

102+191k 91%

```
User: Can you write me a poem about faults and love?  
Assistant: Sure, here's a poem about faults and love:  
They came to tell your faults to me,  
They named them over one by one;  
I laughed aloud when they were done,  
I knew them all so well before,  
Oh, they were blind, too blind to see  
Your faults had made me love you more.
```

Data Preparation

isaacrehg/poetry-instructions

```
conversation
string · lengths
```

```
102+191k    91%
```

```
User: Can you write me a poem about faults and love?
Assistant: Sure, here's a poem about faults and love:
They came to tell your faults to me,
They named them over one by one;
I laughed aloud when they were done,
I knew them all so well before,
Oh, they were blind, too blind to see
Your faults had made me love you more.
```

```
return {
    "prompt": prompt,
    "completion": completion
}
```

```
['prompt', 'completion']
{'prompt': 'Can you write me a poem about faults and love?',
```

```
'completion': "Sure, here's a poem about faults
```

```
DatasetDict({
  train: Dataset({
    features: ['conversation'],
    num_rows: 1764
  })
  validation: Dataset({
    features: ['conversation'],
    num_rows: 111
  })
  test: Dataset({
    features: ['conversation'],
    num_rows: 331
  })
})
```

- We are using model's Tokenizer to encode and decode messages.
- Splitting the dataset
 - 80% used for training
 - 20% used for validation



Training Strategy - *Sequential Fine-Tuning*

1. Domain-Adaptive Pretraining

- Training < 0.3% of parameters
- Focused on poem content
- Repetition Penalty = 1.1
- Temperature = 0.7
- Top_p = 0.95

2. Task Fine-Tuning

- Training 100% of parameters
- General purpose training
- Repetition Penalty = 1.1
- Temperature = 0.7
- Top_p = 0.95

Inspired on approach from BioBERT / ClinicalBERT (Lee et al., 2019)



After Experiment Final Setup

Batch Training

Hyperparameter choice (*gpt/d-s*)

Parameter-Efficient Fine-Tuning (PEFT)

- Use of LoRA (Low-Rank Adaptation)
- Freezing parameters (memory saving and faster training)

```
|| trainable%: 0.2364
```

```
trainable params: 294,912 || all params: 124,734,720 ||
```

Training epochs = 5 | 3

Learning rate = $3e-5$ | $1e-4$

Weight-decay = 0.01 | 0

Device-batch-size = 2 | 2

Gradient Accum Steps = 16 | 4



04 ✨ Results ✨





LLMs Performance Metrics Overview

N-gram Overlap Metrics

Exact Match (EM) / Accuracy

Perplexity

Diversity-n

Human Preference Scores



LLMs Performance Metrics

Training and Validation Error

Using default loss functions of the models.

For GPT2 and Deepseek, it is `torch.nn.CrossEntropyLoss` by default.

$$l_n = -w_{y_n} \log \frac{\exp(x_{n,y_n})}{\sum_{c=1}^C \exp(x_{n,c})}$$

Practical Examples

Both generic and poem explicit prompts will be explored.



Results highlights



DataSet: *isaacrehg/poetry-instructions*

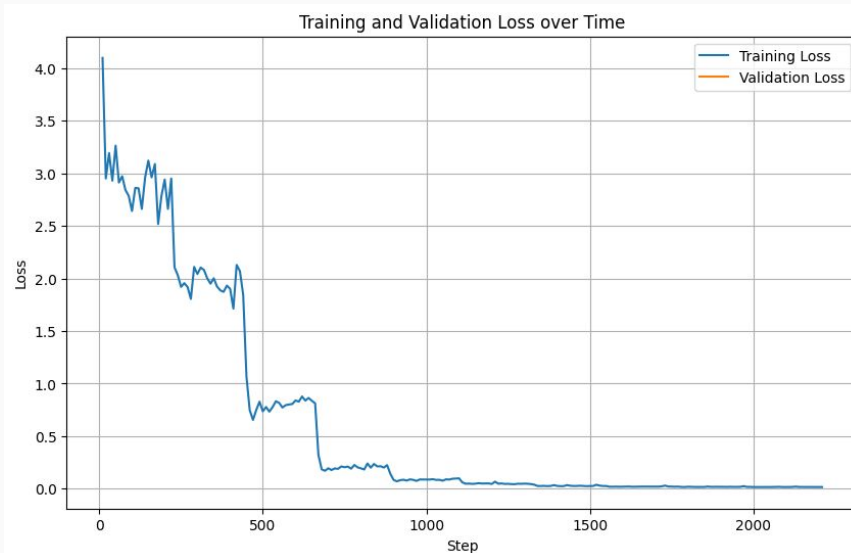
Model: *gpt2*

| per_device_train_batch | gradient_accumulation_steps | num_train_epochs | learning_rate | training loss | val_loss |
|------------------------|-----------------------------|------------------|---------------|---------------|----------|
| 2 | 4 | 1 | 1,00E-04 | 2,3700 | 2,1700 |
| 2 | 8 | 3 | 5,00E-05 | 2,4758 | 2,1564 |
| 2 | 16 | 5 | 3,00E-05 | 2,4570 | 2,1580 |

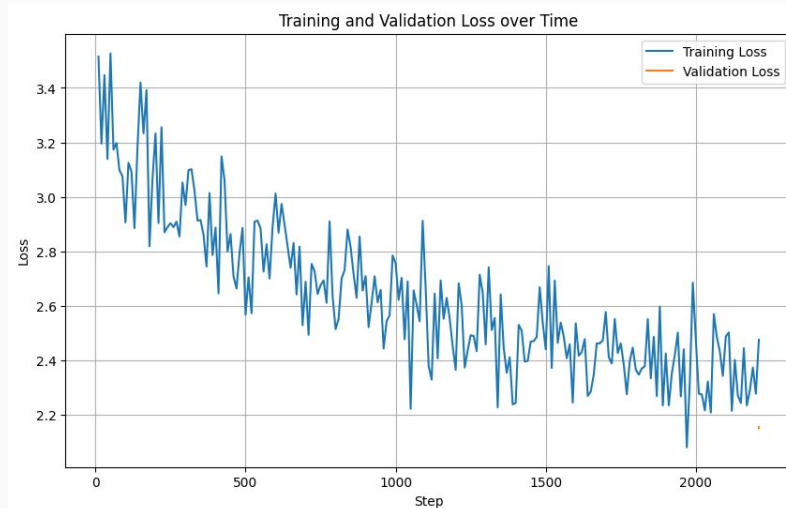
Model: *deepseek-coder-1.3b-base*

| per_device_train_batch | gradient_accumulation_steps | num_train_epochs | learning_rate | training loss | val_loss |
|------------------------|-----------------------------|------------------|---------------|---------------|----------|
| 2 | 4 | 1 | 1,00E-04 | 0,0163 | 0,0157 |
| 4 | 4 | 2 | 2,00E-04 | 0,0163 | 0,0160 |

Best Results



- Training Error: 0.0163
- Validation Error: 0.01575
- Test Error: 0.01577



ChatGPT

- Training Error: 2.4758
- Validation Error: 2.1564
- Test Error: 2.1512



Practical Examples (1/3)

`prompt = "The sea is blue, the sky is red"`

The sea is blue, the sky is red, and the ocean is green.

- The sea is blue, but it's not a real color. It's called "blue" in another language.*
- You can see the blue when you look at it. But to be honest, it isn't really blue. Blue is just like red, except with different wavelengths. (...)*



*The sea is blue, the sky is red.
It's like a good-bye, and the sun goes down.
I don't know what to do with it, but I must go.
You can see it on my screen in the middle of
the night.*



*The sea is blue, the sky is red;
And I hear the great bird of dawn calling his
song. Go and fetch him!
We have come to an end: he has departed
from us.
I will not leave you alone now--let me go
with thee again.
Go and fetch Him again!
Let there be light at last for my soul's end...*

Practical Examples (2/3)

prompt = "The sea is blue, the sky is red"

Setting `pad_token_id` to `eos_token_id`:32014 for open-end generation.

The sea is blue, the sky is red, and the ocean is green.

- The sea is blue, but it's not a real color. It's called "blue" in another language.
- You can see the blue when you look at it. But to be honest, it isn't really blue. Blue is just like red, except with different wavelengths.
- Red is a color that has wavelengths between 400 and 600 nm. This means that if you look at a red light source, you will see a red.
- If you look at the sky, you will see a blue sky. This is because the sky is blue. But the sky is not blue. It's a color.
- There are many colors in nature. There are many different colors. Some colors are more blue than others. Some colors are more red than others.



Setting `pad_token_id` to `eos_token_id`:32014 for open-end generation.

The sea is blue, the sky is red.

It's like a good-bye, and the sun goes down.

I don't know what to do with it, but I must go.

You can see it on my screen in the middle of the night.

And it's not even dark. The sky is red-the sky is red-it's a dark sky, it's not

Is it a dream? Yes, it's a dream,

and you think that dreams are always true. You want to be true. It's a dream.

But I couldn't tell you anything from the night.

When I wake up, I feel like I have nothing left to do.

All I can do is stand there and wait for the sunrise.

Even if I die, I won't get out of here again.

My body doesn't know how to fight. It's a dream.

If I lose my mind, I won't understand why I have to fight.

There's no way out of here.



Setting `pad_token_id` to `eos_token_id`:32014 for open-end generation.

The sea is blue, the sky is red,

The waves are noise, the crickets are sing,

And the little random notes are sweet,

And the busy beaver that weaves the rhyme

Is in itself a real good sense.

A prose is not a prose, but a rhyme.

And a prose may be much too long for a single line,

But the song may be better if it is not for the line.

For the poem is a large subject, yet the smaller parts

Sometimes only seem minor, and you never learn why.

The sea is blue, the sky is red,

And the waves are noise;

The crocuses and daffodils

Are in itself a real good sense.



Practical Examples (3/3)

```
prompt = "Write a poem about the sea and the sky"
```

Write a poem about the sea and the sky.

25 characters or less

Use `print()` to write your code.

"""

Write your code below:

def print_poem():

print("""\nIn a hole in the ground

There lived a king,

Who had a ship

To save his life."

--William Shakespeare

""")



*The sea is full of clouds and mist,
And I am far from cloud and mist.
I am far from the wrangling waves
That sweep the land as they are sails.
I am far from the rushing stream
That shoots the land as it is sails.*

*The sea is full of moaning noise,
And I am far from moaning noise.
I am far from the harp strings note.
(...)
I am far from the minstrels moon
That mocks the lantern on the sea.*



05 ✨ Conclusions

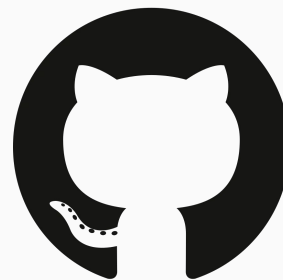
Learnings

- Expensive hardware
- Time-consuming
- Lots of open source and open access tools
- The amount of parameters have a huge impact on the final performance

Individuals can produce notable fine-tuned models with open source tools and limited computational resources.



Hugging Face





Our Experiment Findings

- GPT-2 vs DeepSeek
 - classic
 - modern
 - general purpose
 - code optimized
- LoRA → small fraction of parameters
 - + focused dataset → high quality, domain specific models, with limited resources
- Dataset matters:
 - Overfitting
 - Nonsense outputs



At the end

We obtained a model that:

- Coherent , creative poems
- Diverse structure and vocabulary
- Accurate
- Poetic style
- Generalizes well to unseen prompts



Thanks!



Any question?

Uriel Cabañas
Pau Colomer
Martí Esquius

