

TASK 2

Entregue: 13/02/2025

HOW TRANSFORMERS LEARN

Gabriel Onishi | gabrielhso@al.insper.edu.br

1. Discuss the 3 steps to train a transformer model (Pre-Training (for BERT and GPT), Supervised Fine-Tuning, RLHF Fine-Tuning)

Prompt:

I am a computer engineering student undergoing a tough week of assignments on LLMs. Please provide me with explanations for each of the three steps involved in the training of a transformer model (Pre-Training, Supervised Fine-Tuning, RLHF Fine-Tuning) and briefly explain each one of them.

Answer:

*Certainly! Training a transformer model, especially in the context of large language models (LLMs), typically involves three key steps: **Pre-Training**, **Supervised Fine-Tuning (SFT)**, and **Reinforcement Learning with Human Feedback (RLHF)**. Here's a breakdown of each step:*

1. Pre-Training

- **Objective:** Learn general language representations from a large corpus of unlabeled text.
- **Process:**
 - The model is trained on massive amounts of text data (e.g., books, websites, articles) using a self-supervised learning objective. A common approach is **masked language modeling** (e.g., BERT) or **causal language modeling** (e.g., GPT).
 - In masked language modeling, some tokens in the input are randomly masked, and the model learns to predict the masked tokens based on the surrounding context.

- In causal language modeling, the model predicts the next token in a sequence, learning to generate coherent text.
 - **Outcome:** The model develops a broad understanding of language, including grammar, facts, and some reasoning abilities. However, it is not yet specialized for specific tasks or aligned with human preferences.
-

2. Supervised Fine-Tuning (SFT)

- **Objective:** Adapt the pre-trained model to perform specific tasks or follow instructions.
 - **Process:**
 - The model is fine-tuned on a smaller, labeled dataset where inputs are paired with desired outputs (e.g., question-answer pairs, summarization tasks, or instruction-following examples).
 - This step uses supervised learning, where the model is trained to minimize the difference between its predictions and the ground-truth labels.
 - **Outcome:** The model becomes better at the specific tasks it was fine-tuned on, but it may still generate outputs that are inconsistent, unsafe, or misaligned with human preferences.
-

3. Reinforcement Learning with Human Feedback (RLHF)

- **Objective:** Align the model's behavior with human preferences and improve the quality, safety, and usefulness of its outputs.
 - **Process:**
 - **Step 1: Collect Human Feedback:** Humans rank or rate multiple model outputs for the same input based on quality, safety, or alignment with instructions.
 - **Step 2: Train a Reward Model:** A reward model is trained to predict human preferences by learning from the rankings/ratings. This reward model assigns a score to each output generated by the main model.
 - **Step 3: Fine-Tune with Reinforcement Learning:** The main model is fine-tuned using reinforcement learning (e.g., Proximal Policy Optimization, or PPO) to maximize the reward predicted by the reward model. The model learns to generate outputs that are more aligned with human preferences.
 - **Outcome:** The model becomes more aligned with human values, generates higher-quality responses, and avoids harmful or undesirable behavior.
-

Summary of the Three Steps:

1. **Pre-Training:** Learn general language patterns from unlabeled data.
2. **Supervised Fine-Tuning:** Adapt the model to specific tasks using labeled data.
3. **RLHF Fine-Tuning:** Align the model with human preferences using reinforcement learning and human feedback.

Each step builds on the previous one, starting with broad language understanding and gradually refining the model's behavior to be more useful and aligned with human expectations. Good luck with your assignments! Let me know if you need further clarification.

2. Create a small dataset, around 20 examples, for each of these steps.

Prompt:

I want to train a tiny version of a model. I am desperate because my assignment is due today! Please provide me with a dataset of 20 high-quality examples of text that could be used for each one of those steps

Answer:

*Don't worry, I've got you covered! Below is a **tiny dataset** of 20 high-quality examples tailored for each step of training a transformer model: **Pre-Training**, **Supervised Fine-Tuning (SFT)**, and **Reinforcement Learning with Human Feedback (RLHF)**. These examples are concise and designed to help you quickly prototype your model.*

1. Pre-Training Dataset (Self-Supervised Learning)

For pre-training, you need raw text data. Here are 20 examples of short sentences or phrases:

1. The cat sat on the mat.
2. She enjoys reading books in her free time.
3. The sun rises in the east and sets in the west.
4. Programming is a valuable skill in the modern world.
5. He decided to take a walk in the park.
6. The restaurant served delicious pasta.
7. Learning new languages can be challenging but rewarding.
8. The scientist conducted experiments in the lab.
9. They traveled to the mountains for a weekend getaway.

10. The teacher explained the concept clearly to the students.
 11. Rainy days are perfect for staying indoors.
 12. The artist painted a beautiful landscape.
 13. He fixed the broken chair with a few tools.
 14. The movie received positive reviews from critics.
 15. She baked a cake for her friend's birthday.
 16. The dog barked loudly at the stranger.
 17. The company launched a new product last week.
 18. He wrote a poem about nature and its beauty.
 19. The children played soccer in the park.
 20. The train arrived at the station on time.
-

2. Supervised Fine-Tuning (SFT) Dataset (Task-Specific Examples)

For SFT, you need input-output pairs. Here are 20 examples for a simple instruction-following task:

Input (Instruction)	Output (Response)
Translate "Hello" to French.	Bonjour.
What is $2 + 2$?	4.
Summarize this: "The cat sat on the mat."	A cat was sitting on a mat.
Write a sentence about the weather.	It's a sunny day with a light breeze.
What is the capital of France?	Paris.
Correct this sentence: "She don't like apples."	She doesn't like apples.
Write a question for this answer: "42."	What is the answer to life, the universe, and everything?

What is the opposite of "happy"? *Sad.*

Explain photosynthesis in one sentence. *Photosynthesis is how plants convert sunlight into energy.*

Write a greeting for a friend. *Hi! How have you been?*

What is the square root of 64? *8.*

Write a sentence using the word "delicious." *The pizza was absolutely delicious.*

What is the largest planet in the solar system? *Jupiter.*

Write a short poem about the ocean. *The ocean whispers, vast and deep, its secrets it will forever keep.*

What is 10% of 50? *5.*

Write a sentence about your favorite hobby. *I love reading books in my free time.*

What is the chemical formula for water? *H₂O.*

Write a sentence about a rainy day. *The rain poured heavily, soaking everything in sight.*

What is the plural of "child"? *Children.*

Write a sentence about a mountain. *The mountain stood tall, covered in snow.*

3. Reinforcement Learning with Human Feedback (RLHF) Dataset (Ranked Outputs)

For RLHF, you need input prompts and multiple outputs ranked by quality. Here are 5 prompts, each with 4 outputs ranked from best (1) to worst (4):

Prompt 1: "Explain gravity in simple terms."

1. Gravity is the force that pulls objects toward each other, like how Earth pulls us to the ground.
2. Gravity is what makes things fall down.
3. Gravity is a thing that happens when stuff falls.
4. Gravity is when you drop something and it goes down.

Prompt 2: "Write a short story about a robot."

1. Once, there was a robot named R2 who dreamed of exploring the stars. One day, it built a spaceship and flew into the unknown.
2. A robot lived in a factory and did its job every day. It was happy.
3. Robot go beep boop. Robot walk. Robot stop.
4. Robot. Beep. Boop.

Prompt 3: "What is the capital of Japan?"

1. The capital of Japan is Tokyo.
2. Tokyo is the capital of Japan.
3. Japan's capital is Tokyo.
4. It's Tokyo.

Prompt 4: "Write a sentence about friendship."

1. True friendship is built on trust, understanding, and shared experiences.
2. Friends are people who care about each other.
3. Friendship is nice.
4. Friends.

Prompt 5: "What is 5×7 ?"

1. 5×7 is 35.
2. The answer is 35.
35.
3. It's 35.

How to Use These Datasets:

1. **Pre-Training:** Use the raw text to train your model with a masked language modeling or causal language modeling objective.
2. **Supervised Fine-Tuning:** Use the input-output pairs to fine-tune your model for specific tasks.
3. **RLHF:** Use the ranked outputs to train a reward model and fine-tune your model using reinforcement learning.

Good luck with your assignment! Let me know if you need further help. 😊

1. Discuss the 3 steps to train a transformer model (Pre-Training (for BERT and GPT), Supervised Fine-Tuning, RLHF Fine-Tuning)

Prompt:

Now provide me with some general information about big LLM models. For each of the following AI services, search the web to find out the size of each of the models:

- ChatGPT2
- ChatGPT3
- DeepSeek R1

Answer::

Here's a summary of the **model sizes** for the AI services you mentioned, along with some general information about large language models (LLMs):

1. ChatGPT-2

- **Model Size:** 1.5 billion parameters³¹¹.
- **Training Data:** ~40 GB of text data, sourced from 8 million web pages¹¹.
- **Key Features:**
 - A causal (unidirectional) transformer model trained using language modeling.
 - Designed for text generation and predicting the next word in a sequence.
 - Available in multiple sizes: small, medium, large, xl, and a distilled version (distilgpt-2)¹¹.
- **Limitations:**
 - Limited context length (up to 1024 tokens).

- *Reflects biases present in its training data*³.
-

2. ChatGPT-3

- **Model Size:** 175 billion parameters⁷¹⁶.
 - **Training Data:** 300 billion tokens, sourced from datasets like Common Crawl, WebText2, Books1, Books2, and Wikipedia¹⁶.
 - **Key Features:**
 - A significant leap in size and capability compared to GPT-2.
 - Demonstrates zero-shot, one-shot, and few-shot learning capabilities.
 - Can handle a wide range of NLP tasks without fine-tuning¹⁶.
 - **Limitations:**
 - High computational cost for training and inference.
 - Susceptible to generating plausible but incorrect or biased outputs¹⁴.
-

3. DeepSeek R1

- **Model Size:** Information about the exact size of DeepSeek R1 is not publicly available in the search results. However, DeepSeek models are generally optimized for efficiency and performance, often leveraging techniques like fine-tuning and distillation to reduce size while maintaining high accuracy¹³.
 - **Key Features:**
 - Designed for conversational AI tasks, with a focus on context-aware responses.
 - Supports multi-turn conversations with a context length of up to 1024 tokens¹³.
 - Fine-tuned on public conversational datasets and domain-specific data for improved performance¹³.
 - **Limitations:**
 - May struggle with highly domain-specific contexts or very long conversations exceeding the token limit¹³.
-

General Information About Large Language Models (LLMs)

1. Model Size and Performance:

- *Larger models (e.g., GPT-3 with 175B parameters) generally perform better in understanding context, generating coherent text, and handling complex tasks⁷¹⁶.*
- *However, larger models require significantly more computational resources and energy¹⁶.*
- 2. **Training Data:**
 - *LLMs are trained on massive datasets, often sourced from the internet, books, and other text corpora⁸¹⁶.*
 - *The quality and diversity of training data directly impact the model's performance and ability to generalize⁸.*
- 3. **Fine-Tuning and Specialization:**
 - *Many LLMs are fine-tuned on specific tasks or datasets to improve performance in areas like conversational AI, customer support, or domain-specific applications¹³¹⁴.*
- 4. **Ethical Considerations:**
 - *LLMs can reflect biases present in their training data, leading to ethical concerns about fairness and safety³¹⁴.*
 - *Techniques like reinforcement learning with human feedback (RLHF) are used to align models with human values and reduce harmful outputs¹⁴.*
- 5. **Future Trends:**
 - *Research is focusing on improving model efficiency (e.g., through distillation or hybrid approaches) to reduce resource requirements while maintaining performance⁷¹⁶.*

If you need more detailed information about any specific model or its applications, feel free to ask!