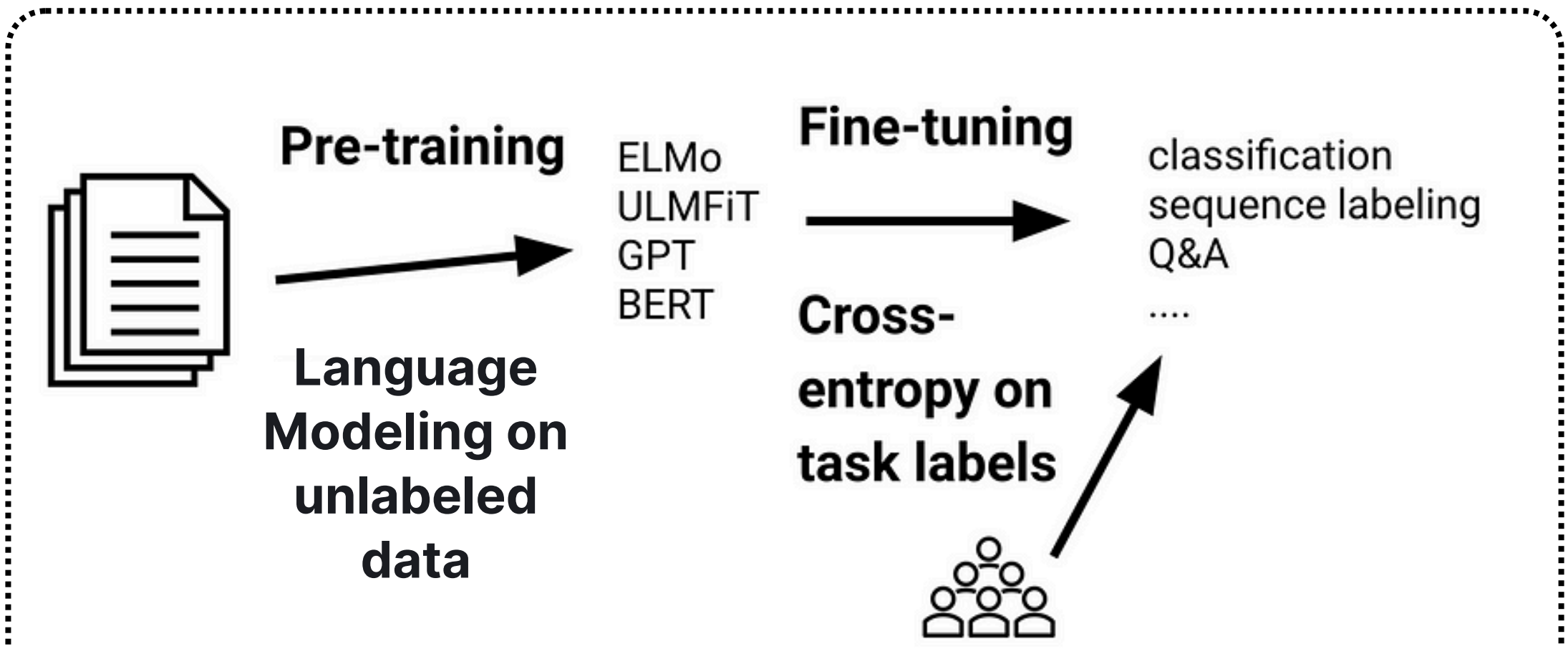
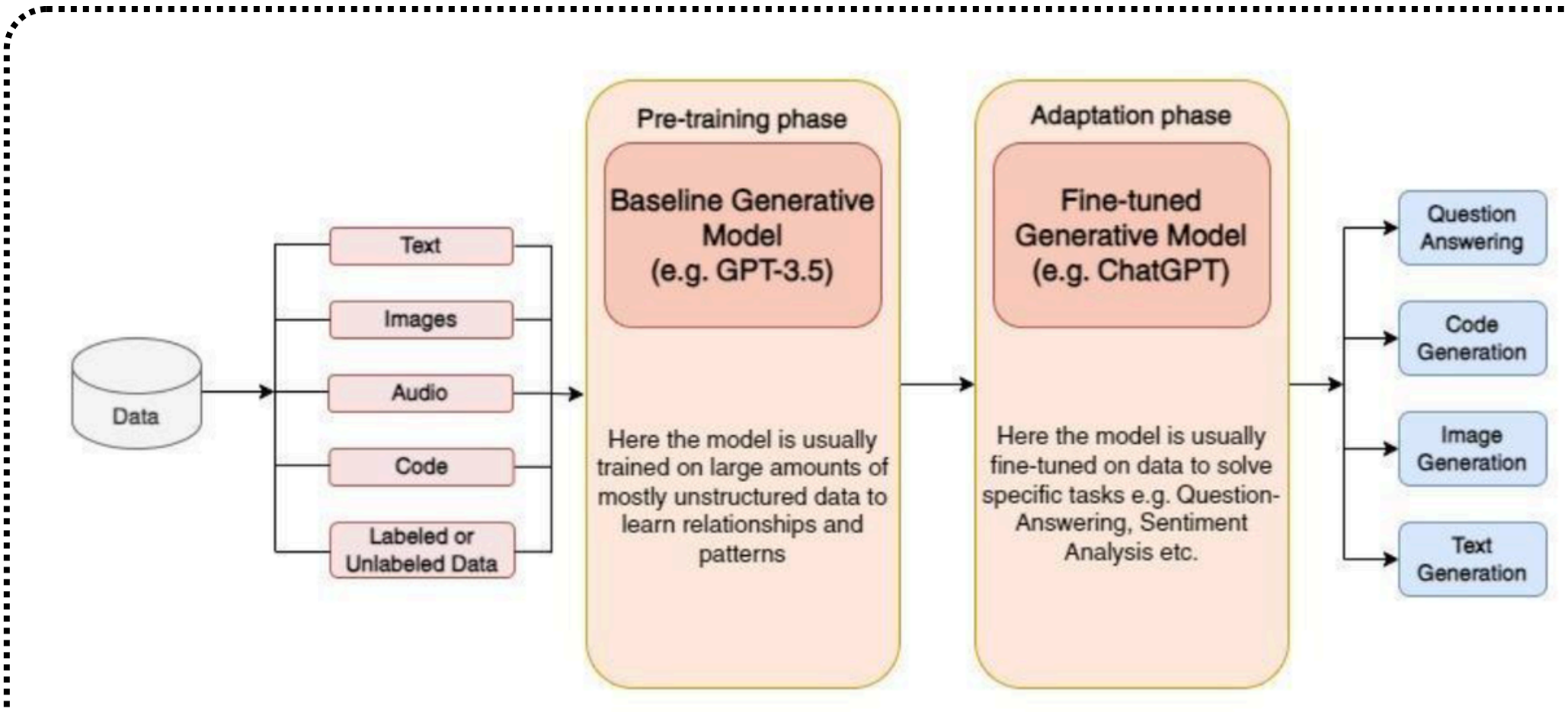
Analytics Vidhya

Mastering LLMs

Day 4: Pre-Training & Fine-Tuning LLMs





The two-phase training (**pre-training** + **fine-tuning**) is a hallmark of transformer-based models and reflects their reliance on large-scale, unstructured data.

Classical ML models, by contrast, adopt a **one-step** training process and directly learn from labeled data, making them simpler but less versatile.

Transformers follows a **two-phase** sequential training approach:

Phase 1
Pretraining

The model is trained on a large corpus of unlabeled data.

Phase 2
Fine-tuning

The pretrained model is fine-tuned on smaller, task-specific datasets.

Phase 1: Pretraining

- In this phase, the model is trained on a **large corpus of unlabeled data**.

- The goal is to learn general-purpose representations that are transferable to a wide variety of downstream tasks.

- Common pretraining tasks include:

    - **Language modeling**: Predicting the next token (e.g., GPT) or missing tokens (e.g., BERT).

    - **Masked Language Modeling (MLM)**: Filling in the blanks for masked words in a sentence.

    - **Next Sentence Prediction (NSP)**: Predicting whether two sentences follow each other logically.

- Models like **GPT**, **BERT**, and **T5** are pretrained on massive datasets such as Common Crawl or Wikipedia.

Phase 2: Fine-tuning

- The pretrained model is fine-tuned on **smaller**, **task-specific** datasets.

- This involves training the model with annotated data specific to the target application (e.g., sentiment analysis, question answering, prompt-responses).

- Fine-tuning adjusts the pretrained representations to fit the nuances of the task.

This **two-phase approach** allows transformers to:

- Leverage large amounts of unlabeled data to build strong initial representations.

- Require fewer labeled samples for fine-tuning, as the model already has a strong foundation.

Classical Machine Learning

Classical ML models usually follow a **one-step training approach**:

- They are trained directly on labeled data relevant to the specific task.

- These models do not separate the learning process into general-purpose pretraining and task-specific fine-tuning.

- Training involves optimizing the model parameters to minimize the error on the training data using techniques like:

  - Minimizing mean squared error for regression.

  - Maximizing likelihood or minimizing cross-entropy for classification.

Transfer Learning as the Foundation

Transfer learning refers to using knowledge gained from solving one task (or learning from one dataset) and applying it to a different but related task.

- It is the overarching concept that unites pre-training and fine-tuning:
    - **Pre-training**: The phase where the model learns general-purpose knowledge from a large, diverse dataset.
    - **Fine-tuning**: The phase where the pretrained model is adapted to specific tasks using smaller, task-specific datasets.

In this context:

- Pretraining is the transfer of general knowledge to a model.
- Fine-tuning customizes and refines that knowledge for particular tasks.

The Workflow

Step 1: Pretraining

- The model is trained on a large corpus of data (e.g., Wikipedia for text, ImageNet for images) to develop general representations.

- **Example in NLP**:
  - A model like BERT is pretrained on billions of sentences using objectives like Masked Language Modeling (MLM).

- **Outcome**:
  - The model develops a strong understanding of language patterns, grammar, syntax, and semantic relationships.

Step 2: Transfer (via the pretrained model)

- The pretrained model is treated as the starting point for solving a new task.

- The representations learned during pretraining are transferred and adapted to the new tasks.

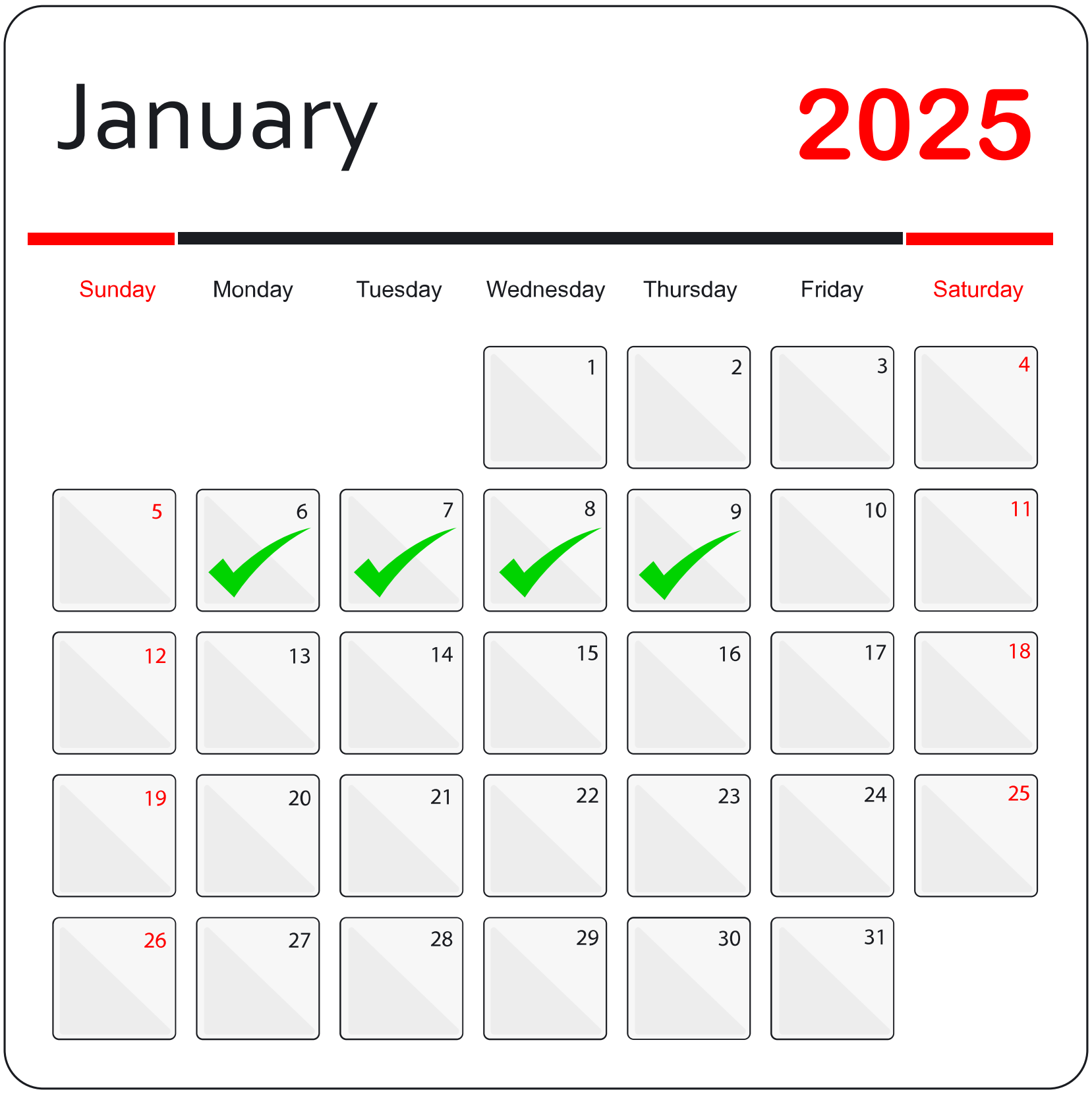
Step 3: Fine-tuning

- Using task-specific labeled data (often much smaller than the pretraining dataset), the model is trained further.

- Fine-tuning updates the model's parameters to align them with the requirements of the specific tasks.

- Example in NLP:
  - Fine-tuning a pretrained BERT model for sentiment analysis on a dataset of movie reviews.
  - Fine-tuning pretrained GPT-4 model into ChatGPT on a dataset of prompts-responses

Analytics Vidhya



Stay Tuned for **Day 5** of

Mastering LLMs