1. Stage One: Pre-training (Building the Foundation)

Pre-training is the process of teaching the model general knowledge by showing it a massive amount of text.

- **The Goal:** To create a "Foundational Model" (or Base Model) that understands language and general facts.
- **The Data:** This stage uses **unlabeled (raw) data** scraped from the internet. For example, GPT-3 was trained on ~300 billion "tokens" (words) from sources like Common Crawl (the entire web), Wikipedia, books, and Reddit links,,.
- **The Task:** The model is given a very simple job: **"Next Word Prediction."** You give it a sentence like *"The lion is in the..."* and it must learn to predict *"forest"*,.
- **The "Magic" Discovery:** The lecture highlights a key realization from OpenAI in 2018: Even though the model is only taught to predict the next word, it unintentionally learns to perform complex tasks like translation, summarization, and sentiment analysis without being explicitly trained for them,.
- **The Cost:** This stage is incredibly expensive. Dr. Dander notes that the pre-training cost for GPT-3 was estimated at **$4.6 million** due to the massive computational power required.

2. Stage Two: Fine-tuning (Specialization)

Fine-tuning is the process of taking that general Foundational Model and refining it for a specific job using a smaller, **labeled dataset**.

- **Why is it needed?** A general model (like a base GPT-4) might be good at everything but master of nothing. It doesn't know private company data, specific legal precedents, or how to speak like a specific customer support agent.
- **The Data:** Unlike pre-training, this stage uses **labeled data**. This comes in two main forms:
    1. **Instruction Fine-tuning:** Giving the model specific "Question-Answer" pairs (e.g., instructions on how to translate specific technical terms).
    2. **Classification Fine-tuning:** Giving the model labeled examples (e.g., showing it 5,000 emails labeled "spam" and 5,000 labeled "not spam").
- **Real-World Examples:**
    - **Legal:** A tool called **Harvey** was fine-tuned specifically on legal case history to help attorneys, something a general model couldn't do reliably,.
    - **Telecom: SK Telecom** fine-tuned a model for Korean customer service, resulting in a 35% improvement in conversation quality compared to the base model.
    - **Banking: JP Morgan** fine-tunes models on their proprietary banking data to assist employees.

Summary Comparison

Feature	Pre-training	Fine-tuning
Objective	Create a "Generalist" (Foundational Model)	Create a "Specialist" (Specific App)
Data Type	**Unlabeled / Raw Text** (e.g., Wikipedia)	**Labeled Data** (e.g., Q&A pairs)
Data Volume	Massive (Billions/Trillions of words)	Smaller, targeted datasets
Cost	Extremely High ($Millions)	Significantly lower

The "Sky is the Limit" Workflow

The lecture concludes by outlining the 3-step workflow for building an industrial-grade AI,:

1. **Data Collection:** Gather raw text.
2. **Pre-training:** Train a Foundational Model (Expensive, High Compute).
3. **Fine-tuning:** Train on Labeled Data (Targeted, Specific Application).
