

# METHODOLOGY

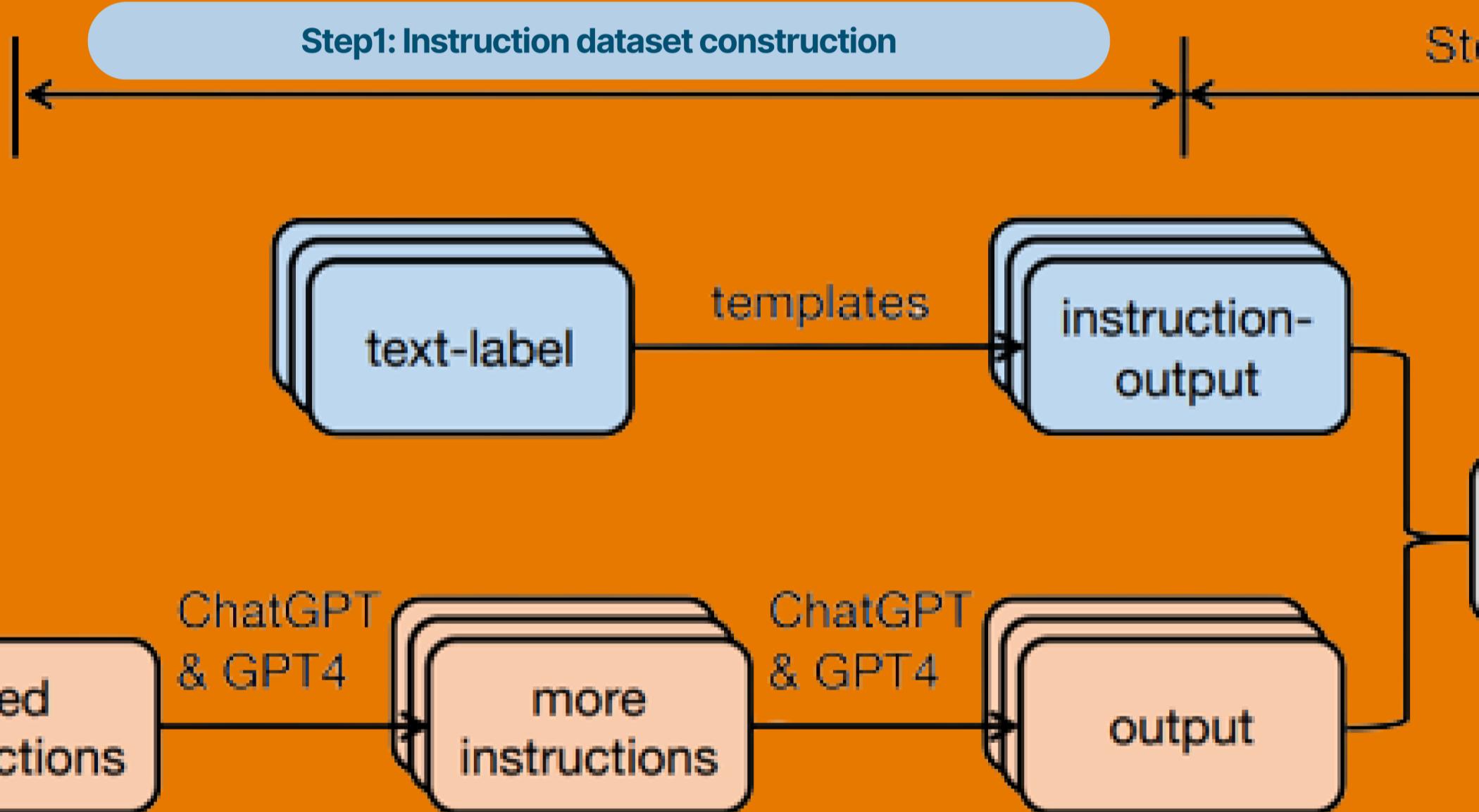
## Instruction finetuning Pipeline



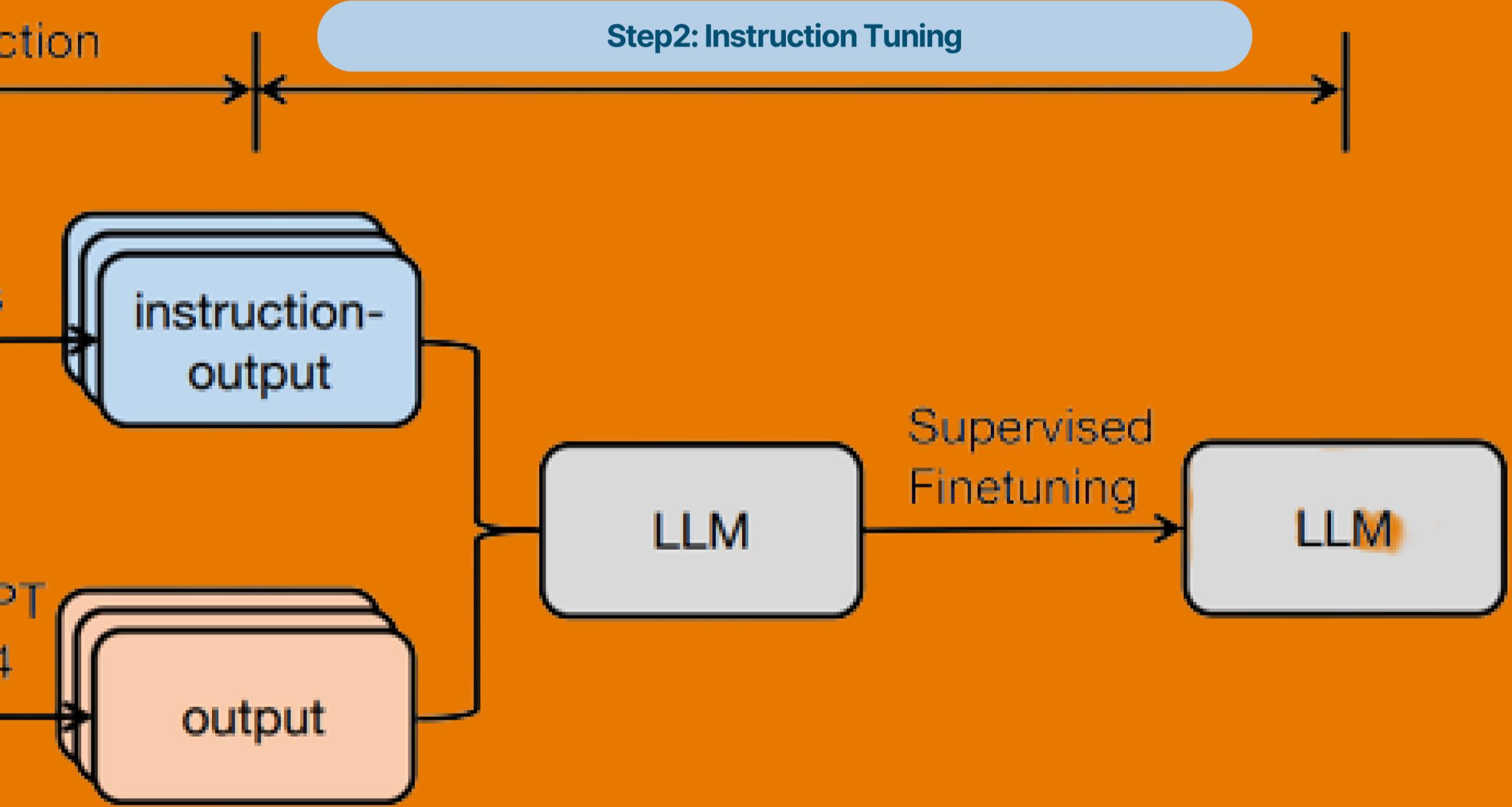
Ashish Patel  
@ashishpatel2604



# INSTRUCTION DATASET CONSTRUCTION



# INSTRUCTION TUNING



## Step 01

### CRAFTING INSTANCES

1. **Instruction:** Natural language text defining the task (e.g., write a thank-you letter, create a blog post).
2. **Optional Input:** Extra info for context.
3. **Anticipated Output:** Expected result from the instruction and input.



## Step 02

# DATASET CREATION

Two methods for dataset construction:

### ✳️ Data Integration:

Gather (instruction, output) pairs from existing annotated datasets, using templates for transformation.

### 🚀 LLMs for Output:

Use language models like GPT-3.5-Turbo to generate outputs, either from manual or seed instructions. Example datasets: Flan, P3.



## Step 03

# BUILDING DATASETS

Datasets like **InstructWild** and **Self-Instruct**:

- Constructed using LLM-generated outputs.
- Instructions from collections or expanded using LLMs.
- Great for quick dataset creation.



## Step 04

# CONVERSATION IT DATASETS

For multi-turn conversations:

-  Large models play different roles.
-  Generate messages in conversation format.
- Example: User and AI assistant interactions.



# Step 05

## TUNING PRETRAINED MODELS

Using the collected IT dataset:

-  Direct fine-tuning in a fully-supervised manner.
-  Predict each token sequentially.
- Train the model on the instruction and input.



# Step 06

## CONVERSATION IT DATASETS

For multi-turn conversations:

-  Large models play different roles.
-  Generate messages in conversation format.
- Example: User and AI assistant interactions.



# Step 07

## IN SUMMARY

1. Create instruction dataset with instruction, optional input, and anticipated output.
2. Use data integration or LLMs to build datasets efficiently.
3. Consider conversational datasets for dynamic interactions.
4. Fine-tune models based on IT dataset for accurate results.

Explore the power of instruction-based fine-tuning!



# I'm Ashish Patel

I am Data Science Expert and Researcher at IBM

I am LinkedIn Top Data Science Voice Helping the people to guide on their Career path to achieve their goals

Follow me and together, we'll spend days making a learning easy



**Ashish Patel**  
@ashishpatel2604

