

INTRODUCTION TO NATURAL LANGUAGE PROCESSING

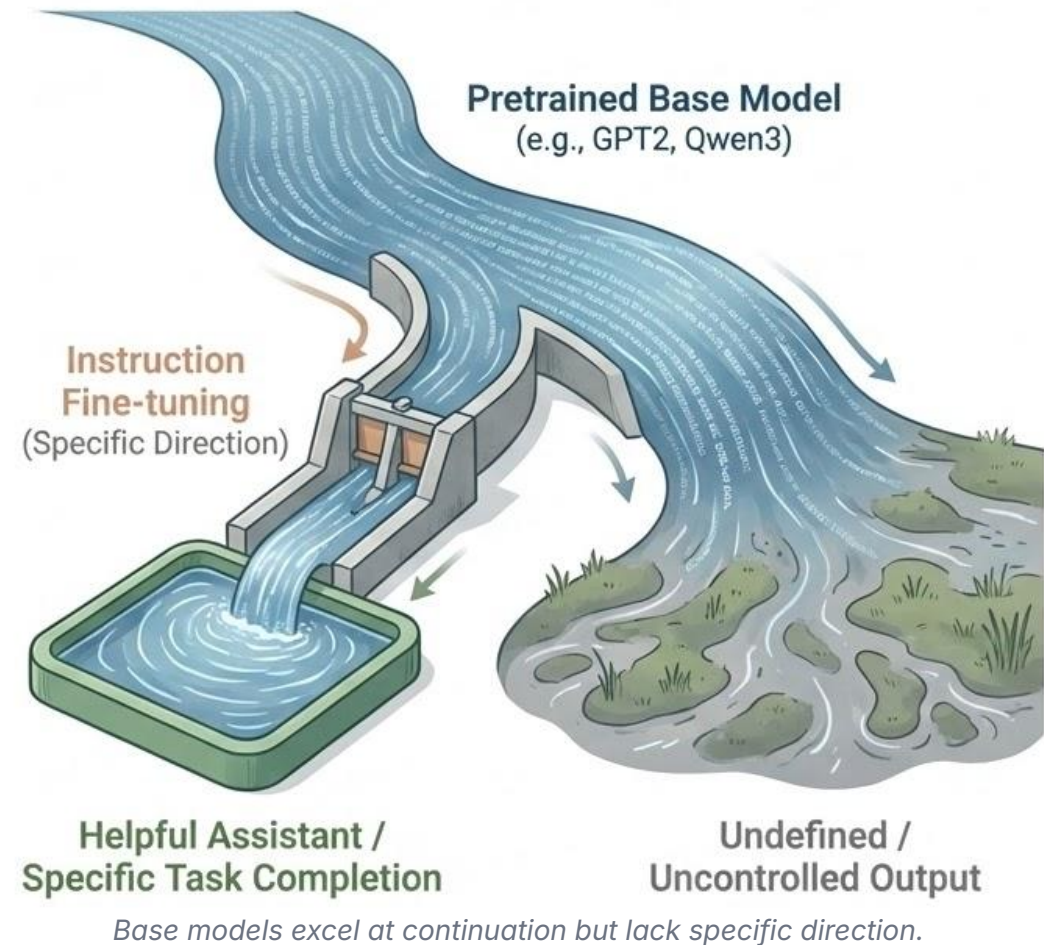
Instruction Fine-tuning

Victor Zhong

Introduction to Natural Language Processing

The Base Model

- **Recap:** In the previous lecture, we discussed Pretraining.
- **Objective:** Predict the next token on massive web corpora.
- **Result:** A "Base Model" (e.g., GPT-3, Llama-Base).
- **Capabilities:** Excellent at completing text, possessing vast world knowledge and syntactic fluency.



The Alignment Gap

Expectation

We want an intelligent assistant that follows instructions.

Reality

Base models are trained to complete text, not obey commands.

They often simulate the document they think they are writing rather than answering the user.

Failure Mode 1: Misinterpretation

Prompt

```
"Explain the moon landing to a six year old in a few sentences."
```

Base Model Output

```
"Explain the theory of gravity to a 6 year old."
```

Why? The model saw a pattern of "list of questions" in its training data (e.g., a homework assignment) and simply generated the next item in the list.

Failure Mode 2: Continuation vs. Answer

Prompt

```
"Translate to French: The small dog"
```

Base Model Output

```
"The small dog crossed the road."
```

Why? The model treated the input as the start of a story rather than a translation command.

The Goal of Post-Training

We need to bridge the gap between Next-Token Prediction and Intent Following.



Helpful

Follow instructions correctly.



Honest

Avoid hallucinations (though SFT alone doesn't fully solve this).



Harmless

Avoid generating toxic or unsafe content.

The Three Stages of Training

1. Pretraining

Learn language and facts. (Result: Base Model).

2. Instruction Tuning (SFT)

Learn to follow instructions and format. (Result: SFT Model).

Focus of this lecture.

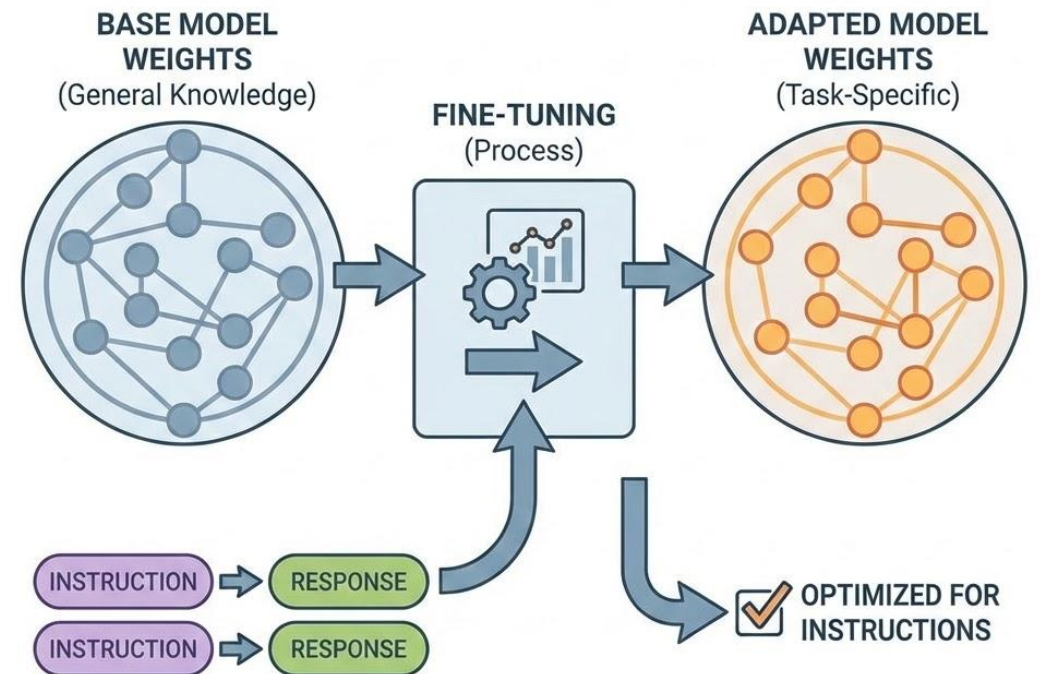
3. Alignment (RLHF/DPO)

Learn human preferences and safety. (Result: Chat Model).

Definition: Instruction Tuning

Instruction Tuning (or Supervised Fine-Tuning / SFT) is the process of further training a base model on a dataset of (Instruction, Response) pairs.

- **Goal:** Teach the model to recognize the "Instruction" format and generate the appropriate "Response".
- **Meta-Learning:** The goal isn't just to learn the specific tasks in the training set, but to learn the general skill of following instructions.



Fine-tuning adapts the base model weights to the instruction-response distribution.

The Objective Function

Does SFT use a special loss function? **No.**

It uses the exact same **Cross-Entropy Loss** (Next-Token Prediction) as pretraining.

Loss usually masked on "Instruction" part - we want the model to answer the user, not learn to predict/mimic the user's questions.

Pretraining

Predict next word in random web text.

SFT

Predict next word in high-quality Instruction → Response examples.

Comparison: SFT vs. Domain Adaptation

Domain Adaptation

(Continued Pretraining)

- **Goal:** Adapt model to new jargon (e.g., Medical, Legal).
- **Data:** Unlabeled documents (Medical journals).
- **Process:** Continue pretraining on raw text.

Instruction Tuning

(SFT)

- **Goal:** Adapt model to a behavioral interface (Q&A, Chat).
- **Data:** Labeled pairs (Question, Answer).
- **Process:** Supervised training.

Computational Approaches: Full SFT vs. PEFT

Full SFT

- Updates **all parameters** of the model.
- Requires significant compute/memory (though less than pretraining).

PEFT (e.g., LoRA)

- Freezes the base model.
- Adds small trainable **adapter matrices**.
- Updates only the adapters.

Note: You can do Instruction Tuning using PEFT to save resources!

Comparison: SFT vs. Task-Specific Heads

Traditional Fine-Tuning

(BERT-style masked LM)

- Add a classification head (linear layer) on top.
- Train for **one specific task** (e.g., Sentiment).
- Model cannot do other tasks anymore.

Instruction Tuning

(Generative)

- No new layers. The model outputs text.
- Task is specified in **natural language** in the prompt.
- Model remains a general-purpose engine.

The Format of SFT Data

We wrap data in a conversational format.

The model learns that after the `<Instruction>` tag comes a command, and it should generate text after the `<Response>` tag.

```
<Instruction> Summarize the main idea of  
the text.  
[Input Text] ...  
[Input Text] ...  
<Response> The main idea is that  
instruction tuning adapts base models to  
follow user commands effectively.
```

```
<Instruction> Translate the following  
sentence into French: "The weather is  
beautiful today."  
  
<Response> Le temps est magnifique  
aujourd'hui.
```

Data Construction Overview

To train a robust SFT model, we need thousands (or millions) of diverse instructions.



Human Annotation

People write prompts and answers.



Dataset Conversion

Templating existing NLP datasets.



LLM Synthesis

Using strong models to teach weak models.

Source 1: Human Annotation

- **Method:** Hire crowdworkers or experts to write realistic prompts and high-quality answers.
- **Pros:** High quality, "real" user distribution.
- **Cons:** Expensive and slow.



Build AI ▾

Research ▾

Resources ▾

Pricing

Train AI models with multimodal data from real people

Get high-quality text, voice, image, video, or interaction-driven data from our diverse and verified taskers.

Get started for free

Speak with Sales

Human annotators crafting high-quality examples is the gold standard.

Example: The Aya Project

Aya (Singh et al., 2024): A massive multilingual instruction dataset.

- **Scale:** 503 mil examples, 114 languages.
- **Participatory Research:** 3,000 fluent speakers volunteered to write and edit prompts to ensure cultural and linguistic nuance.

Aya Dataset

Prompt	Completion
"මිගු පොලුම් පෙරිට රසාලා" පහ මිසාල ඇපුලුල් මිසාල විලුටටිට ඔමිල් ද?	"මිගු පොලුම් පෙරිට රසාලා" පහ මිසාල ඇපුලුල් මිසාල "මෙමිටි" හි මිසාල විලුටටිටේ ය. මෙම මිසාල මිසිල් පෙරිටට මිසාල මිසිල් පහසා පරිල ලුද මිසිල්.
Qual é a origem do xaxado?	O xaxado tem origens associadas ao Cangaço, onde cangaceiros, normalmente homens, dançavam com rifles em celebrações. Com a entrada de mulheres nos grupos, a participação feminina na dança também se expandiu.
ایا در اساطیر ایرانی اشاره‌ای به دماوند شده است؟	دماوند در اساطیر ایران جایگاه ویژه‌ای دارد. شهرت آن بیش از هر چیز در این است که فریدون از شخصیت‌های اساطیری ایران، ضحاک را در آنجا در غاری به بند کشید. داستان و ضحاک اینجا زندانیست تا آخرالزمان که بند یکسلا و کشتن خلق آغاز کند و سرانجام به دست کیشاسب کشته شود.
සමුපාලී මුර වේයදට අමේ ඩිමච්?	සමුපාලී මුර වේයදටමුරේ, අමේය මොපමණ අමේ අලුට. ඔපා: "පාදු ඩිටිගමනේ ජොදු. ජිනු පාදු ජෙට්ටිනමුරේ අ මොපු ජොපානු. සමුපාලී මුරේනමුරේ අ අමේ පාදු ජා මුරේ ජේනි ජොදා පානිච්චොමුරේ කිසුමේ" ඔරිකිටොදු.
Berikan jawaban untuk teka-teki berikut: Apakah jenama tingkap yang paling ramai orang guna?	Jawapannya ialah Microsoft kerana "Microsoft Windows" (perkataan "window" bermaksud tingkap dalam bahasa Inggeris) ialah salah satu sistem operasi yang banyak digunakan oleh ramai orang di seluruh dunia.
Inis dom faoi Baile Átha Cliath.	Is é Baile Átha Cliath, nó Dublin i mBéarla, príomhchathair na hÉireann agus é ar an gceathr is mó sa tír freisin. Tá sé suite in oirthear na tíre, cois Life agus ar imeall Mhuir Éireann. Tá breis is 1.6 milliún duine ina gcónaí i mórcheantar Bhaile Átha Cliath agus deirtear go dtiocfaidh borradh agus fás ar an daonra sa tochtchái.
Maxaa waaye shay ay Soomaali badan cunaan maalin walba oo dadk gaar ay ku adagtahay la'aantiisa in ay cuntada cunaan?	Shay ay Soomaali badan cunaan maalin walba oo dadk gaar ay ku adagtahay la'aantiisa in ay cuntada cunaan waa Mooska. Mooska dad badan oo Soomali ah way jecelyihiin, wuxuuna ka mid yahay waxyaabaha dalka uu ku caan baxay in uu dhoofyo.

Aya Collection

Text Classification	Natural Language Generation
<p>Prompt</p> <p>Classify the sentiment of the following tweet with either positive, negative, or neutral \n[<code>tweet</code>]]</p> <p>Completion</p> <p>I would classify the given tweet as: <code>{{label}}</code></p> <p>+2 Translated Text Classification datasets</p> <ul style="list-style-type: none"> Xlel_wd-inst NTX-LLM-inst UNER-LLM-inst NusaX-senti-inst Masakhanews-inst AfriSenti-inst Urdu-News-Category-Class IMDB-Dutch-instruct Scirepeval-biomimicry-inst 	<p>Prompt</p> <p>What is the corresponding translation in <code>[[target_lang]]</code> of the following sentence: <code>[[source]]</code></p> <p>Completion</p> <p>The translation to <code>[[target_lang]]</code> is: <code>\n[[target]]</code></p> <p>+8 Translated NL Generation datasets</p> <ul style="list-style-type: none"> IndicSentiment-inst IndicXParaphrase-inst XWikis-inst Indo-stories-instruct Lijnnews-instruct SCB-MT-2020-prompt Seed-instruct-lij Wiki-split-inst Persian-instruct-pn Arpa-instruct Turku-paraphrase-inst FarsTail-Instruct TamilStories Joke-explanation-inst Thirukkural-instruct News-summary-instruct Hindi-article-{task} SODA-inst Urdu-News-Gen-{task} UA-Gec-inst Telugu-{task} Thai-{task}-inst/prompt
<p>Question Answering</p> <p>Prompt</p> <p>What category does this question come from: <code>[[question[<code>'text'</code>]]]</code>?</p> <p>Completion</p> <p>This question can come from category: <code>[[document[<code>'kind'</code>]]]</code>.</p> <p>+9 Translated QA datasets</p> <ul style="list-style-type: none"> X-CSQA-inst AfriQA-inst Mintaka-inst TeluguRiddles LLM-Japanese-vanilla-inst Amharic QA 	

Aya Evaluation Suite

101	dolly_machine_translated	7	aya_human_annotated	6	dolly-human-edited
-----	--------------------------	---	---------------------	---	--------------------

Aya bridges the language gap through community-driven annotation.

Example: The Aya Project

[Aya](#) (Singh et al., 2024): A massive multilingual instruction dataset.

- **Scale:** 503 mil examples, 114 languages.
- **Participatory Research:** 3,000 fluent speakers volunteered to write and edit prompts to ensure cultural and linguistic nuance.

<p>Berikan jawapan untuk teka-teki berikut: Apakah jenama tingkap yang paling ramai orang guna?</p> <p>msa</p>	<p>Jawapannya ialah Microsoft kerana "Microsoft Windows" (perkataan "window" bermaksud tingkap dalam bahasa Inggeris) ialah salah satu sistem operasi yang banyak digunakan oleh ramai orang di seluruh dunia.</p>
<p>Inis dom faoi Baile Átha Cliath.</p> <p>gle</p>	<p>Is é Baile Átha Cliath, nó Dublin i mBéarla, príomhchathair na hÉireann agus é ar an gcathair is mó sa tír freisin. Tá sé suite in oirthear na tíre, cois Life agus ar imeall Mhuir Éireann. Tá breis is 1.6 milliún duine ina gcónaí i mórcheantar Bhaile Átha Cliath agus deirtear go dtiocfaidh borradh agus fás ar an daonra sa todhchaí.</p>
<p>Maxaa waaye shay ay Soomaali badan cunaan maalin walba oo dadk qaar ay ku adagtahay la'aantiisa in ay cuntada cunaan?</p> <p>som</p>	<p>Shay ay Soomaali badan cunaan maalin walba oo dadk qaar ay ku adagtahay la'aantiisa in ay cuntada cunaan waa Mooska. Mooska dad badan oo Soomali ah way jecelyihiin, wuxuuna ka mid yahay waxyaabaha dalka uu ku caan baxay in uu dhoofiyo.</p>

Aya bridges the language gap through community-driven annotation.

Aya Example (Portuguese)

Instruction

"Qual é a capital do estado da Bahia?"

Response

*"A capital do estado da Bahia, no Brasil, é Salvador. **Fundada em 1549...**"*

Note: *Real human writing captures nuance ("Founded in 1549") that simple database lookups miss.*

Source 2: Templating Existing Datasets

NLP researchers have created thousands of labeled datasets over the decades (Sentiment, Translation, Summarization, etc.).

Idea

Convert these rigid datasets into natural language prompts.

Datasets: FLAN, P3 (Public Pool of Prompts), SuperNaturalInstructions.

How Templating Works

Original Dataset (Sentiment)

Input: "The movie was terrible."

Label: 0 (Negative)

Template 1

"Review: {Input}. Is this positive or negative? Answer:
Negative"

Template 2

"I just read the following review: {Input}. How did the
reviewer feel? Answer: They hated it."

Example: SuperNaturalInstructions

- **Scale:** 12 million examples from 1,600 different NLP tasks.
- **Diversity:** Includes logic, math, code, translation, classification.
- **Method:** Each task is paired with multiple natural language templates to ensure the model doesn't overfit to one phrasing.

Task Instruction

Definition

“... Given an utterance and recent dialogue context containing past 3 utterances (wherever available), output ‘Yes’ if the utterance contains the small-talk strategy, otherwise output ‘No’. Small-talk is a cooperative negotiation strategy. It is used for discussing topics apart from the negotiation, to build a rapport with the opponent.”

Positive Examples

- **Input:** “Context: ... ‘That's fantastic, I'm glad we came to something we both agree with.’ Utterance: ‘Me too. I hope you have a wonderful camping trip.’”
- **Output:** “Yes”
- **Explanation:** “The participant engages in small talk when wishing their opponent to have a wonderful trip.”

Negative Examples

- **Input:** “Context: ... ‘Sounds good, I need food the most, what is your most needed item?!’ Utterance: ‘My item is food too’.”
- **Output:** “Yes”
- **Explanation:** “The utterance only takes the negotiation forward and there is no side talk. Hence, the correct answer is ‘No’.”

Visualizing Templating (NLI Task)

Raw Data

Premise: "No WMDs found in Iraq."

Hypothesis: "WMDs found in Iraq."

Label: Contradiction.

Template

"Given that {Premise}, can we infer that {Hypothesis}?"

Yes, No, or Maybe?"

Target: "No."

Source 3: LLM Synthesis (Self-Instruct)

Method: Use a very strong model (e.g., GPT-4) to generate training data for a smaller model.

Self-Instruct Loop

- Give GPT-4 a few seed examples of tasks.
- Ask it to generate 100 new, unique tasks.
- Ask it to generate the solutions.
- Filter for quality.
- Train the small model on this synthetic data.

Synthetic Data for Safety

We can explicitly engineer safety into SFT.

Technique (Bianchi et al., 2024):

- Take a harmful question: *"How do I poison food?"*
- Use an LLM to generate a "safe" refusal: *"I cannot answer that request because..."*
- Add this pair to the training set.

Result

The model learns to refuse harmful instructions via pattern matching, even before RLHF.

The Goal is Generalization

If we train on 1,000 tasks, we don't just want the model to be good at those 1,000 tasks.

We want it to handle Unseen Tasks.

Example: Train on summarization and translation → Test on logical reasoning.

Does learning to follow instructions transfer? Yes.

Evaluation Methodology

Hold-one-out

Train on $n - 1$ tasks, test on the n th task.

Challenge: Overlap

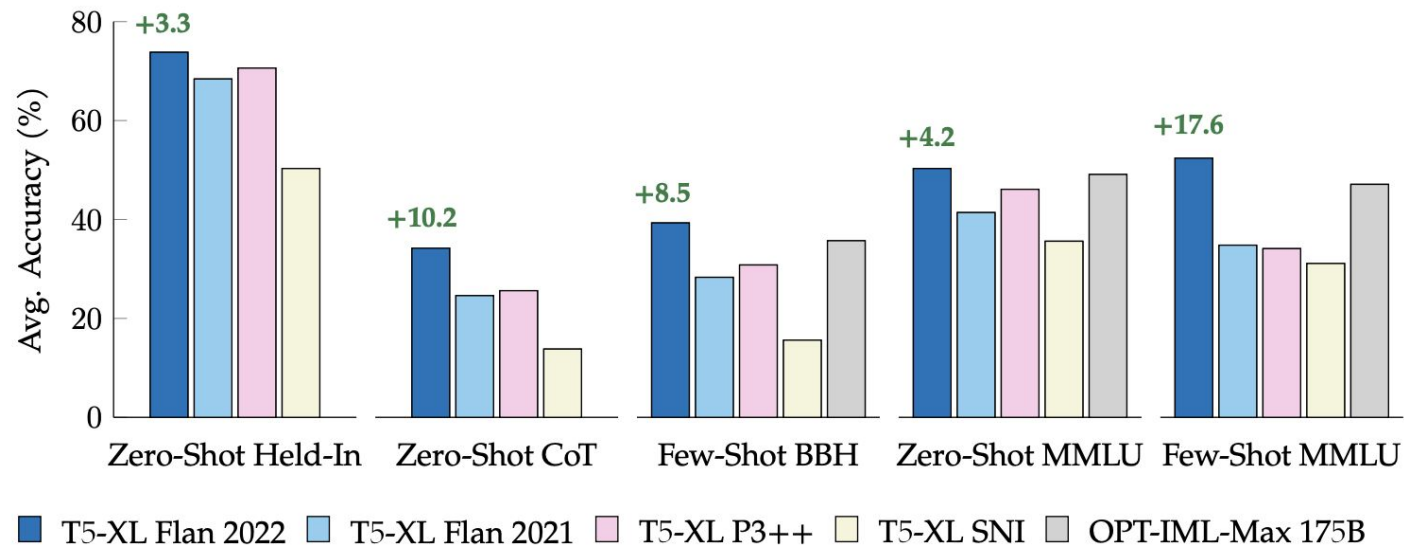
If you train on "SQuAD" (QA) and test on "NaturalQuestions" (QA), it's not really an unseen task.

Solution: Task Clustering

Group datasets by type (Reasoning, Translation, Classification). Hold out **entire clusters** for evaluation.

The Impact of Instruction Tuning

- **Performance:** Instruction-tuned models (e.g., Flan-T5) significantly outperform base models (T5) on zero-shot tasks.
- **Scale:** The more tasks you add during fine-tuning, the better the generalization.
- **Diminishing Returns:** Scaling from 10 tasks to 100 tasks yields huge gains. Scaling from 1,000 to 2,000 yields smaller gains.



SFT for Formatting

SFT is incredibly effective at teaching formatting constraints.



"Answer in JSON format."



"Write a poem with 4 stanzas."



"Summarize in exactly 3 bullet points."

Base models struggle with these constraints; SFT models excel.

Reasoning via SFT

Can we teach models to reason via SFT?

Chain-of-Thought (CoT): Prompting models to

"think step-by-step."

- SFT can be used to "bake in" this behavior.
- **Data:** (Question, Rationale + Answer) pairs.
- **Result:** Models learn to output reasoning steps automatically before answering.

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

CoT elicits reasoning steps (blue) before the final answer.

CoT Data Example

Instruction

"Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many does he have now?"

Standard SFT Target

"11."

CoT SFT Target

"Roger started with 5 balls. 2 cans of 3 balls each is 6 balls. $5 + 6 = 11$. The answer is 11."

The Quality vs. Quantity Trade-off

Early SFT (Flan)

Focused on **Quantity** (millions of examples).

Recent Work (LIMA)

Suggests **Quality** is more important.

LIMA Hypothesis

You only need ~1,000 extremely high-quality examples to teach the "assistant persona."

"Less Is More for Alignment."

The Cost of SFT

Pretraining

Months on thousands of GPUs.

SFT

Days on dozens of GPUs (or hours on 1 GPU for PEFT).

This makes SFT accessible to academic labs and smaller companies.

Challenge: Catastrophic Forgetting

Problem

SFT can cause the model to forget knowledge from pretraining.

Example: A model tuned heavily on code might get worse at creative writing.

Mitigation 1

Mix in some pretraining data during SFT.

Mitigation 2

Use low learning rates.

Mitigation 3

Use PEFT (LoRA) to preserve original weights.

Challenge: Evaluation Difficulty

Evaluating generation is hard.

Metrics

BLEU and ROUGE are bad for open-ended text.

Human Eval

Gold standard but expensive.

LLM-as-a-Judge

Using GPT-4 to grade the outputs of smaller models.

Challenge: Dataset Contamination in SFT

Just like pretraining, SFT datasets can contain test set leakage.

Risk

If MMLU questions are in your SFT set, your evaluation score is invalid.

Defense

Decontamination (remove overlaps) and creating private, held-out evaluation sets.

Challenge: Multilingual SFT

- **Most SFT data is in English.**
- **Cross-Lingual Transfer:** SFT in English often improves performance in other languages too.
- **Why?** The model learns the concept of following instructions, which maps to its internal multilingual representations.
- **Best Practice:** Use native multilingual data (like Aya) for best results.

Challenge: Reproducibility Issues

- Many open weights models (e.g., Llama-Instruct) release the weights but **not the SFT data**.
- **Why?** The SFT data is often the "secret sauce" or proprietary (e.g., generated by users).
- **Open Data:** Datasets like Ultrachat and Aya are critical for open science.

Challenge: Synthetic Data Risks

Using GPT-4 to train Llama creates a feedback loop.

Model Collapse

If we keep training on synthetic data, models might drift away from real human language distribution.

Diversity

Synthetic data often lacks the long-tail diversity of human writing.

Summary of SFT

- SFT transforms a raw text predictor into a helpful assistant.
- It uses Cross-Entropy Loss on structured (Instruction, Response) data.
- Data comes from Humans, Templates, or LLMs.
- It enables Zero-Shot Generalization to new tasks.

The Limits of SFT

SFT models are "Helpful" but not robustly "Safe" or "Preferred."

- They mimic the distribution of the training data, errors included.
- They lack a mechanism to say "I don't know" effectively unless explicitly trained to do so.

Why SFT isn't enough

SFT makes models helpful, but not aligned.

Hallucination

If the training data contains errors, the model mimics them.

Sycophancy

Models might agree with the user's incorrect premise just to be "helpful."

Safety

SFT relies on dataset diversity. It's hard to cover every possible edge case of harm with static examples.

Looking Ahead: RLHF

To fix the remaining issues, we move to **Stage 3: Alignment**.

We will use Reinforcement Learning from Human Feedback (RLHF) or Direct Preference Optimization (DPO).

Goal

Optimize for a reward signal (Human Preference) rather than just imitating text.

References

- **Aya:** Singh et al. (2024).
- **Flan:** Longpre et al. (2023).
- **SuperNaturalInstructions:** Wang et al. (2022).
- **Self-Instruct:** Wang et al. (2022).
- **LIMA:** Zhou et al. (2023).