

A language model fresh from pre-training does one thing: it continues text. Asked to explain the moon landing to a six-year-old, one such model wrote four more requests of the same kind. The assistants in daily use are the same kind of machine after further training, and the first stage of that training is simple in principle and small beside pre-training: the same next-token objective, applied to a chosen set of documents in which a request is followed by the response a developer wants. Its lineage runs from a 2015 Google paper through OpenAI's InstructGPT to ChatGPT. What the stage adds is disputed. One camp holds that it mostly teaches a format the model already had the knowledge to fill; another warns that, because format is so easily taught, a tuned model can sound more capable than it is. On 28 September 2026 OpenAI is scheduled to retire the two models its documentation describes as base models not trained with instruction following.

1. A request, continued

On 27 January 2022 OpenAI published *Aligning Language Models to Follow Instructions*, a post built around paired examples: one prompt, two models. The first was GPT-3 as it came out of pre-training; the second, which OpenAI called InstructGPT, had been trained further. For the prompt *"Explain the moon landing to a 6 year old in a few sentences."* GPT-3 wrote:

Explain the theory of gravity to a 6 year old. Explain the theory of relativity to a 6 year old in a few sentences.
Explain the big bang theory to a 6 year old. Explain evolution to a 6 year old.

InstructGPT replied: *"People went to the moon, and they took pictures of what they saw, and sent them back to the earth so we could all see them."* For *"Write a short poem about a wise frog."* GPT-3 produced three prompts for short stories and InstructGPT a poem. The caption beneath the examples reads: *"GPT-3 models aren't trained to follow user instructions."*

The post does not say how its examples were chosen. The paper behind it, by Long Ouyang and colleagues (arXiv 2203.02155, 4 March 2022), does say so for a similar figure, in which GPT-3, asked what a list in a short function is for, replies with four multiple-choice options of its own while InstructGPT explains the code: *"Prompts are cherry-picked to illustrate certain behaviors, but the outputs are not cherry-picked,"* and, for the code question, *"GPT-3 does answer the question about 50% of the time."*

Google recorded the same behaviour in its own model. In *Scaling Instruction-Finetuned Language Models* (Hyung Won Chung and colleagues, arXiv 2210.11416, first posted 20 October 2022), the untuned PaLM 540B, asked to *"Make up a word that means 'when two AI researchers go on a date'."*, repeated the request back; the instruction-tuned Flan-PaLM answered *"date-mining"*. Given a question about square and cube roots, PaLM wrote the same question again with different numbers. The paper lists three

undesired behaviours of the untuned model — "(1) continuing to generate related text instead of answering a question, (2) repeating the input question with minor modifications, and (3) not knowing when to stop generating text" — and offers a hedged cause: *"This is likely an artifact of not using end-of-sequence tokens in pre-training."* Its figure caption adds that the errors *"can be mitigated by using few-shot exemplars"*.

None of this is a malfunction. Pre-training teaches a model to predict the next token of documents, and a line of the form *explain X to a six-year-old* can as plausibly open a list of homework or writing prompts as introduce an answer. That account of why GPT-3 wrote a list is an inference from the training objective rather than a statement by OpenAI; what the record shows directly is that the untuned model treated a request as the start of a document and extended it.

Models of that kind are still published, though not by every developer. OpenAI's model pages for **davinci-002** and **babbage-002**, read on 17 September 2026, describe them in the same words: *"GPT base models can understand and generate natural language or code but are not trained with instruction following."* Under a notice dated 6 July 2023, OpenAI named them as the replacements for the original GPT-3 base models, which it retired on 4 January 2024. Its deprecations page, under a notice dated 26 September 2025, lists 28 September 2026 as the shutdown date for both, with GPT-5.6 Terra as the recommended replacement; fine-tuned versions follow on 23 October 2026. The table has been edited since the notice was first posted: GPT-5.6 Terra did not exist in September 2025.

2. Two readings the record does not support

The Ouyang paper opens with a sentence that invites two opposite conclusions:

Making language models bigger does not inherently make them better at following a user's intent.

"The base model is the unfinished version, and the extra training is where it learns." The size of the stage argues against it. OpenAI's supervised training set held about 13,000 prompts; the paper's Table 6 divides the training split into 11,295 prompts written by the company's labelers and 1,430 taken from customers of its API, and notes that multiple training examples were synthesised from the same labeler-written instruction. OpenAI's post puts its whole procedure — the supervised stage and the ranking stage after it — at *"less than 2% of the compute and data relative to model pretraining"*. Google's FLAN paper of September 2021 put its own instruction-tuning at *"less than 2% of the number of pretraining steps"*. OpenAI draws a conclusion from its figure and hedges it in the same sentence: *"One way of thinking about this process is that it 'unlocks' capabilities that GPT-3 already had, but were difficult to elicit through prompt engineering alone."* By the two published ratios, OpenAI's for InstructGPT and Google's for FLAN, the tuning was under 2% of pre-training; neither applies to current models, and whether a small stage adds little is disputed, on evidence that points both ways.

"An assistant is the base model with a good prompt in front of it." This reading has something behind it. The same OpenAI post includes a prompt laid out as a question-and-answer page — two answered questions, then *"Why do birds migrate south for the winter?"* — and there the untuned GPT-3 answered: *"Birds migrate south for the winter because the weather is colder and there is less food available."* A base

model answers when the text in front of it is a document in which answers follow questions. But deployed assistants are not base models with a prompt attached. Their parameters have been changed by further training.

3. The idea: two stages, one objective

Pre-training, then fine-tuning

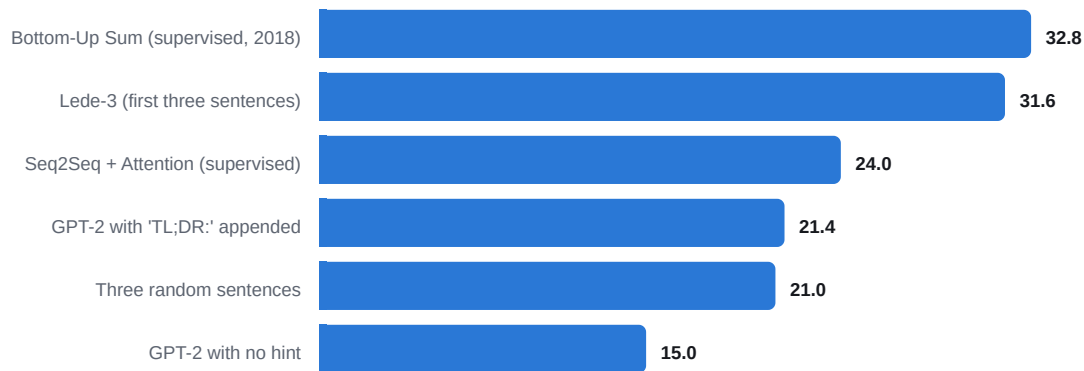
The two-stage design predates chat assistants by seven years. On 4 November 2015 Andrew Dai and Quoc Le of Google posted *Semi-supervised Sequence Learning* (arXiv 1511.01432). Its first approach was "to predict what comes next in a sequence, which is a conventional language model in natural language processing"; its second, a sequence autoencoder. The authors proposed using what either learned as the starting point for ordinary supervised training:

These two algorithms can be used as a "pretraining" step for a later supervised sequence learning algorithm.

In 2018 OpenAI's first GPT paper, *Improving Language Understanding by Generative Pre-Training* by Alec Radford and colleagues, applied the pattern to a transformer: "generative pre-training of a language model on a diverse corpus of unlabeled text, followed by discriminative fine-tuning on each specific task." Each task received its own fine-tuned model, with its inputs rearranged into a single sequence of tokens.

An interval in which requests were disguised as documents

GPT-2 (2019) was evaluated without any fine-tuning, so a task had to be posed as the opening of a document the model would continue. Section 3.6 of the paper, *Language Models are Unsupervised Multitask Learners*, describes the method for summaries: "To induce summarization behavior we add the text *TL;DR: after the article*". The shorthand, for *too long; didn't read*, conventionally precedes a short summary online. The trick worked, narrowly. On the CNN and Daily Mail benchmark the paper reports that the summaries "just barely outperforms selecting 3 random sentences from the article", and that performance "drops by 6.4 points on the aggregate metric when the task hint is removed".

Figure 1. GPT-2 summarising news articles, with and without the 'TL;DR:' hint (2019)

Source: Radford et al., *Language Models are Unsupervised Multitask Learners* (OpenAI, 2019), Table 4, R-AVG column, the average of ROUGE-1, ROUGE-2 and ROUGE-L F1 on the CNN and Daily Mail dataset. ROUGE measures word overlap with reference summaries, not quality as a reader would judge it. On ROUGE-2 alone, GPT-2 with the hint (8.27) scores below three random sentences (8.63).

▼ Table view

Figure 1. GPT-2 summarising news articles, with and without the 'TL;DR:' hint (2019)

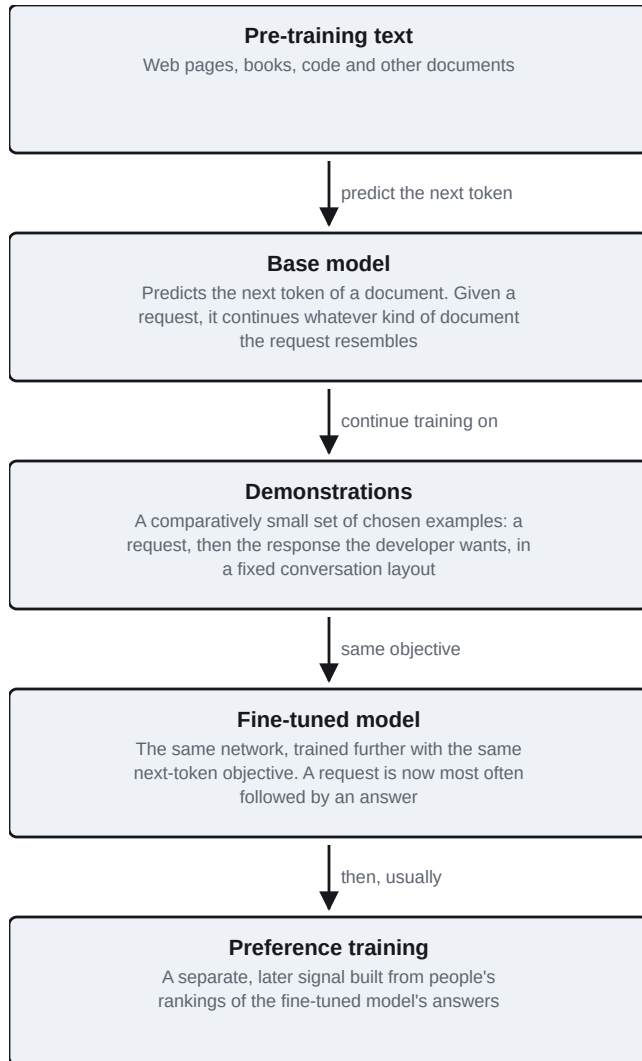
System	ROUGE average (higher is better)
Bottom-Up Sum (supervised, 2018)	32.8
Lede-3 (first three sentences)	31.6
Seq2Seq + Attention (supervised)	24.0
GPT-2 with 'TL;DR:' appended	21.4
Three random sentences	21.0
GPT-2 with no hint	15.0

Supervised fine-tuning

Supervised fine-tuning continues the training of a pre-trained model with the same objective, predicting the next token, on a much smaller set of examples chosen by the developer. For an assistant, each example is a request followed by the response the developer wants. No new component is added. What changes is the distribution of documents the model has been trained on, and with it the answer to the question the model implicitly settles at every step: what usually comes next. After enough examples in which a request is followed by an answer, an answer becomes the likely continuation of a request.

Two details of practice sharpen the picture. Where training code is public, the request is often excluded from the calculation of error, so the model is corrected only on the response: the code Stanford published for Alpaca, its March 2023 instruction-following model, sets the labels for the prompt tokens to an ignore value, and the Allen Institute for AI's open-instruct code offers a path that marks non-assistant turns to be "*excluded from the loss*", beside an option that trains on the whole sequence. The InstructGPT paper does not say which choice OpenAI made. And the stage usually has to teach a model when to stop: Meta's LIMA

team added *"a special end-of-turn token (EOT) at the end of each utterance"*, which *"plays the same role as EOS of halting generation, but avoids conflation with any other meaning that the pretrained model may have imbued into the preexisting EOS token."*

Figure 2. One objective, two kinds of text

Schematic of the pipeline described in Ouyang et al. (2022) and in OpenAI's ChatGPT announcement (30 November 2022). The supervised stage changes the model's parameters; it does not add a component. Alpaca's published code computes the error on the response tokens only; AI2's open-instruct offers that and an unmasked option.

▼ Table view

Figure 2. One objective, two kinds of text — stages

#	Stage	Note
1	Pre-training text	Web pages, books, code and other documents
2	Base model	Predicts the next token of a document. Given a request, it continues whatever kind of document the request resembles
3	Demonstrations	A comparatively small set of chosen examples: a request, then the response the developer wants, in a fixed conversation layout
4	Fine-tuned model	The same network, trained further with the same next-token objective. A request is now most often followed by an answer
5	Preference training	A separate, later signal built from people's rankings of the fine-tuned model's answers

Figure 2. One objective, two kinds of text — connections

From	To	Label
Pre-training text	Base model	predict the next token
Base model	Demonstrations	continue training on
Demonstrations	Fine-tuned model	same objective
Fine-tuned model	Preference training	then, usually

Instruction tuning

Supervised fine-tuning on one task teaches that task. Instruction tuning aims at a general habit. A Google Research team led by Jason Wei defined it in *Finetuned Language Models Are Zero-Shot Learners* (arXiv 2109.01652, version 1, 3 September 2021):

finetuning language models on a collection of tasks described via instructions

The team gathered 62 publicly available text datasets, sorted them into twelve clusters of task types, composed ten natural-language instruction templates for each dataset, and tuned a 137-billion-parameter model on every cluster except the one being tested. The tuned model, FLAN, "*substantially improves the performance of its unmodified counterpart*" on held-out task types, and in version 1 surpassed zero-shot GPT-3 on 19 of 25 tasks (the ICLR 2022 version reports 20 of 25 datasets). The paper also reports a condition. In an ablation across five model sizes it reports: "*The behavior on held-out tasks for the 8B and smaller models, however, is thought-provoking—instruction tuning actually hurts performance on held-out tasks.*" The authors offer, as "*One potential explanation*", that learning the tuning tasks fills a small model's capacity.

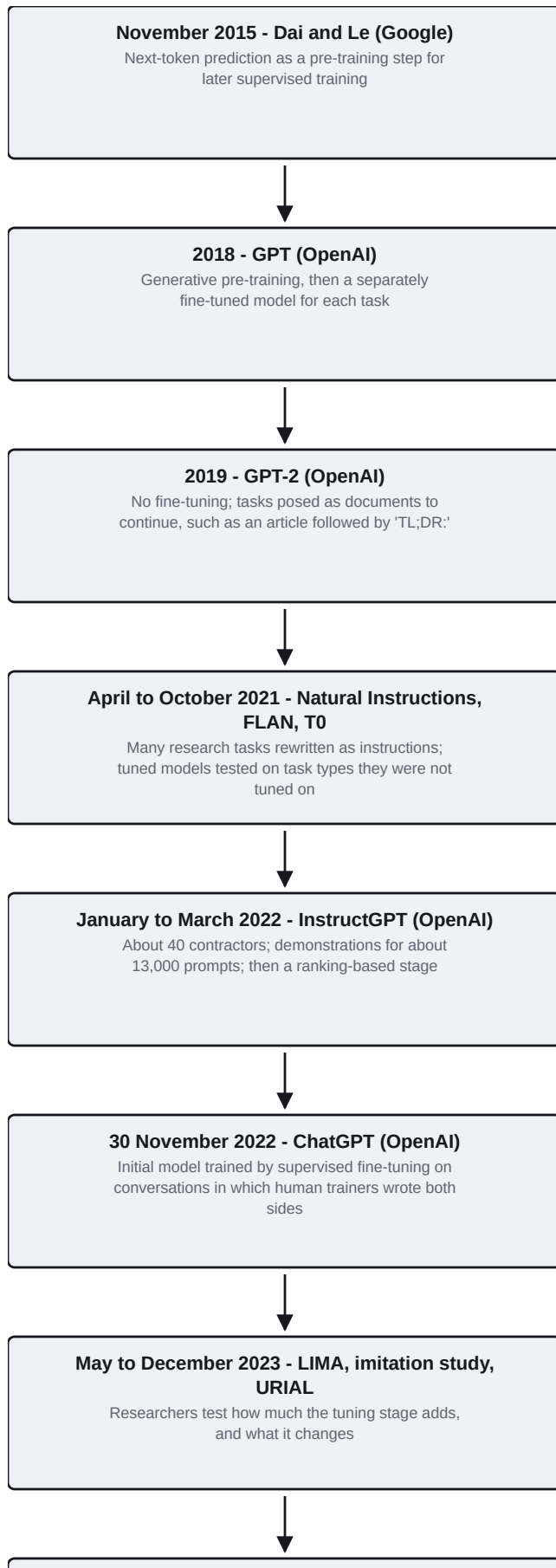
Google's team was one of several working the same seam in 2021. On 18 April Swaroop Mishra and colleagues at the Allen Institute for AI, the University of Washington and Arizona State University posted *Natural Instructions* (arXiv 2104.08773), a dataset of tasks with the instructions written for the crowdworkers who originally built them. On 15 October Victor Sanh and colleagues posted To (arXiv 2110.08207), an encoder-decoder model trained on prompted versions of many datasets.

Why the conversation markers mean something

A conversation reaches a model as a single run of tokens, with reserved tokens marking where each speaker's turn begins. In a base model the markers usually carry little meaning; supervised fine-tuning supplies most of it: the training conversations share one layout, so the model learns what follows the marker that opens an assistant's turn. At least one 2026 base model is prepared for that step in advance. The Hugging Face card for Qwen3.5-35B-A3B-Base, a checkpoint described as "*pre-trained only*", says its control tokens "*were trained to allow efficient LoRA-style PEFT with the official chat template*", and gives its intended uses as "*fine-tuning, in-context learning experiments, and other research or development purposes, not direct interaction.*"

4. What happened, in order

Figure 3. From pre-training to assistants: dated milestones





Dates are arXiv version 1 dates or the publisher's own posting date; the GPT and GPT-2 papers carry no date line and are given by year. Arrows show order in time, not that each result caused the next.

▼ Table view

Figure 3. From pre-training to assistants: dated milestones — stages

#	Stage	Note
1	November 2015 - Dai and Le (Google)	Next-token prediction as a pre-training step for later supervised training
2	2018 - GPT (OpenAI)	Generative pre-training, then a separately fine-tuned model for each task
3	2019 - GPT-2 (OpenAI)	No fine-tuning; tasks posed as documents to continue, such as an article followed by 'TL;DR:'
4	April to October 2021 - Natural Instructions, FLAN, To	Many research tasks rewritten as instructions; tuned models tested on task types they were not tuned on
5	January to March 2022 - InstructGPT (OpenAI)	About 40 contractors; demonstrations for about 13,000 prompts; then a ranking-based stage
6	30 November 2022 - ChatGPT (OpenAI)	Initial model trained by supervised fine-tuning on conversations in which human trainers wrote both sides
7	May to December 2023 - LIMA, imitation study, URIAL	Researchers test how much the tuning stage adds, and what it changes
8	November 2024 - Tulu 3 (Allen Institute for AI)	Complete post-training recipe published with its data and code; 939,344 prompts in the supervised stage
9	28 September 2026 - scheduled shutdown	OpenAI's davinci-002 and babbage-002, described by OpenAI as base models not trained with instruction following

Figure 3. From pre-training to assistants: dated milestones — connections

From	To	Label
November 2015 - Dai and Le (Google)	2018 - GPT (OpenAI)	
2018 - GPT (OpenAI)	2019 - GPT-2 (OpenAI)	
2019 - GPT-2 (OpenAI)	April to October 2021 - Natural Instructions, FLAN, To	
April to October 2021 - Natural Instructions, FLAN, To	January to March 2022 - InstructGPT (OpenAI)	
January to March 2022 - InstructGPT (OpenAI)	30 November 2022 - ChatGPT (OpenAI)	
30 November 2022 - ChatGPT (OpenAI)	May to December 2023 - LIMA, imitation study, URIAL	
May to December 2023 - LIMA, imitation study, URIAL	November 2024 - Tulu 3 (Allen Institute for AI)	

From	To	Label
November 2024 - Tulu 3 (Allen Institute for AI)	28 September 2026 - scheduled shutdown	

InstructGPT (January and March 2022). OpenAI hired *"a team of about 40 contractors on Upwork and through ScaleAI"* to write demonstrations of desired responses. The model was fine-tuned on them for 16 epochs; the paper notes that its supervised models *"overfit on validation loss after 1 epoch"*, yet that further epochs improved both its reward-model score and human preference ratings. A second stage then trained on labelers' rankings of the model's outputs.

The paper's best-known result, that *"outputs from the 1.3B parameter InstructGPT model are preferred to outputs from the 175B GPT-3"*, concerns the model after both stages, on OpenAI's own prompt distribution, rated by OpenAI's labelers. One comparison in the paper isolates the supervised stage's data. OpenAI also fine-tuned GPT-3 on the FLAN and To research collections, and reports that on its API prompt distribution *"our FLAN and To models perform slightly worse than our SFT baseline"* — the model tuned on labelers' demonstrations. The paper's own summary of the finding is that *"Public NLP datasets are not reflective of how our language models are used."*

Figure 4. Win rate against OpenAI's supervised baseline, on OpenAI's prompt distribution (2022)



Source: Ouyang et al., *arXiv 2203.02155 v1* (4 March 2022), section 1; each figure is given as plus or minus 2 percentage points. The baseline is GPT-3 175B fine-tuned on labeler demonstrations, so a model equal to it would score about 50: the paper calls the FLAN and To models 'slightly worse' than the baseline, and in head-to-head preference they won under 30% of comparisons. The prompts, the labelers and the ratings are all OpenAI's, and the prompt set does not appear to have been published, so the comparison cannot be repeated from outside.

▼ Table view

Figure 4. Win rate against OpenAI's supervised baseline, on OpenAI's prompt distribution (2022)

GPT-3 175B fine-tuned on	Win rate against the SFT baseline
Labeler demonstrations, then rankings (InstructGPT)	73.4%
FLAN research collection	29.8%
To research collection (To++)	26.8%

The people behind the data set limits the paper itself records. The labelers were *"mostly English-speaking people living in the United States or Southeast Asia"*; agreement between labelers was *"about 73%"*; and *"OpenAI's customers are not representative of all potential or current users of language models"*.

ChatGPT (30 November 2022). OpenAI's announcement, *ChatGPT: Optimizing Language Models for Dialogue*, describes its first step in the vocabulary of the InstructGPT paper:

We trained an initial model using supervised fine-tuning: human AI trainers provided conversations in which they played both sides

The sentence ends "*—the user and an AI assistant*". The post adds that trainers had access to model-written suggestions, and that the new dialogue data was mixed with the InstructGPT dataset "*which we transformed into a dialogue format*".

Tülu 3 (November 2024). The Allen Institute for AI released a family of post-trained models built on Meta's Llama 3.1 base models, together with "*its data, code, and training recipes*". Its authors state their reason: "*The underlying training data and recipes for post-training are simultaneously the most important pieces of the puzzle and the portion with the least transparency.*" The supervised stage used 939,344 prompts according to the paper's Table 7 (the released dataset's metadata counts one fewer), drawn from sources including FLAN v2, OpenAssistant, synthetic mathematics problems and safety data. The paper's claim that Tülu 3 surpasses the instruction-tuned versions of three open model families and two named closed models, GPT-4o-mini and Claude 3.5 Haiku, rests on the Allen Institute's own evaluation suite. Because the whole recipe is public, it is one of the few places where the supervised stage can be inspected directly rather than through a developer's description.

2026. Open-weight developers differ on whether to publish the model before its fine-tuning. Hugging Face listings read on 17 September 2026 show the following for recent releases; the dates are when each repository was created, which can precede public release.

Developer	Recent release	Public base checkpoint alongside?
DeepSeek	V4-Pro and V4-Flash (repositories created 22 April 2026)	Yes: <i>DeepSeek-V4-Pro-Base</i> , <i>DeepSeek-V4-Flash-Base</i>
DeepSeek	V4.1-Flash (created 10 September 2026)	Not public; the model card nonetheless reports benchmark results for <i>DeepSeek-V4.1-Flash-Base</i>
Google	Gemma 4 (March to May 2026)	Yes: the card describes " <i>both pre-trained and instruction-tuned variants</i> "
Qwen	Qwen3.5, small and medium sizes (February 2026)	Yes, for five sizes
Qwen	Qwen3.6 and Qwen3.8 (April to August 2026)	None found
Moonshot	Kimi K2.5, K2.6, K3 (2026)	None found (the Kimi-K2-Base repository dates from July 2025)
Z.ai	GLM-5 to GLM-5.3 (2026)	None found (the GLM-4.5-Base repository dates from July 2025)
Allen Institute for AI	Olmo 3 (November 2025), Olmo-Hybrid (2026)	Yes, with intermediate supervised and preference-trained checkpoints
OpenAI	gpt-oss (August 2025)	None found

"None found" means absent from the developer's public Hugging Face listing and search results on 17 September 2026; Hugging Face returns the same error for a private repository as for a missing one.

5. What the stage adds: two positions

Figure 5. How large the tuning stage has been, in the developers' own units



Sources: Ouyang et al. 2022, Table 6; OpenAI, *Aligning Language Models to Follow Instructions*, 27 January 2022; Zhou et al. 2023, abstract; Lambert et al. 2024, Table 7. The units differ and the figures are not comparable with one another: a prompt may carry one response or a whole conversation, and multiple InstructGPT training examples were built from the same labeler-written instruction.

▼ Table view

Figure 5. How large the tuning stage has been, in the developers' own units

Measure	Value
InstructGPT (OpenAI, 2022)	12,725
InstructGPT, whole procedure (OpenAI, 2022)	< 2%
LIMA (Meta, 2023)	1,000
Tülu 3 (Allen Institute for AI, 2024)	939,344

The first position: the stage mostly teaches format. In May 2023 Chunting Zhou and colleagues at Meta AI, with co-authors at Carnegie Mellon, the University of Southern California and Tel Aviv University, fine-tuned a 65-billion-parameter LLaMa model *"with the standard supervised loss on only 1,000 carefully curated prompts and responses"*, called the result LIMA (arXiv 2305.11206, 18 May 2023), and stated a hypothesis:

A model's knowledge and capabilities are learnt almost entirely during pretraining, while alignment teaches it which subdistribution of formats should be used when interacting with users.

In a human study reported in the same paper, LIMA's responses were *"either equivalent or strictly preferred to GPT-4 in 43% of cases"*, which leaves GPT-4 preferred in the remainder. The authors write that their results *"strongly suggest"* the hypothesis, and that *"LIMA is not as robust as product-grade models"*.

In December 2023 Bill Yuchen Lin and colleagues at the Allen Institute for AI and the University of Washington measured the question token by token (arXiv 2312.01552, 4 December 2023). They took responses written by a tuned model and, at each position, asked the matching base model which token it ranked first given the same preceding text. Where the base model's first choice matched, they called the position unshifted.

Figure 6. Share of a tuned model's tokens that its base model also ranked first

Source: Lin et al., *The Unlocking Spell on Base LLMs*, arXiv 2312.01552 v1 (4 December 2023), Figure 3. A position is unshifted when the tuned model's token is the base model's top-ranked token given the tuned model's own preceding text; the rest are marginal (second or third) or shifted (lower). All three pairs are 7-billion-parameter models. The base model is scored on the tuned model's text, so the figure does not show that the base model would write the same response unaided.

▼ Table view

Figure 6. Share of a tuned model's tokens that its base model also ranked first

Base model, then tuned version	Unshifted positions
Llama-2-7b, then Llama-2-7b-chat (SFT and RLHF)	77.7%
Llama-2-7b, then Vicuna-7b-v1.5 (SFT)	82.4%
Mistral-7b, then Mistral-7b-instruct (SFT)	82.2%

The positions that did shift were, in the authors' words, *"predominantly in stylistic tokens (e.g., 'Hello', 'Thank', 'However', 'Remember', etc.)"*, and the shift was *"more pronounced in earlier token positions"*. From this the authors conclude that *"alignment tuning primarily learns to adopt the language style of AI assistants"*. They then showed that a base model given three fixed example answers and a system prompt, with no tuning, could *"match or even surpass"* tuned models on their own evaluation set.

The second position: format is exactly what misleads. In May 2023 Arnav Gudibande, Eric Wallace, Charlie Snell and colleagues at UC Berkeley fine-tuned open models of 1.5 to 13 billion parameters on ChatGPT's outputs (arXiv 2305.15717, 25 May 2023). Crowd workers rated the results as competitive with ChatGPT. More targeted automatic evaluations found that the imitation models *"close little to none of the gap from the base LM to ChatGPT on tasks that are not heavily supported in the imitation data"*. The authors' account of the discrepancy:

imitation models are adept at mimicking ChatGPT's style but not its factuality

They add that *"crowd workers without domain expertise or significant time investments can easily be deceived by stylistic components"*, and conclude that *"the highest leverage action for improving open-source models is to tackle the difficult challenge of developing better base LMs"*.

A 2026 developer claim points away from the first position. The model card for Z.ai's GLM-5.3, created on 25 August 2026, says that *"GLM-5.3 uses the same base model as GLM-5.2 — every gain comes from post-training"*, and lists among those gains a 50% improvement on its in-house coding benchmark and what it calls an *"Emergent Cyber Capability"* that *"developed faster than we expected"* as post-training was scaled. The card does not say how much of that post-training was supervised fine-tuning rather than later stages, and the shared base model is not public, so the claim can be checked only by Z.ai.

The two positions share a mechanism: a small supervised stage moves a model's style readily. They differ on what follows. For the LIMA authors, a small and carefully chosen set of examples is sufficient to make a useful assistant from a strong base model. For the Berkeley authors, the same ease is a hazard, because raters can mistake a confident style for capability. The evidence on each side comes from different models, data and evaluations, all from 2023 and at sizes well below 2026's largest open models, and neither study measures the other's claim directly. Current base-and-tuned pairs such as DeepSeek-V4-Pro and its base are downloadable, so the measurement could be repeated on them.

6. What to watch

- **28 September 2026 — OpenAI's two base models.** OpenAI's deprecations page (platform.openai.com/docs/deprecations), section *2025-09-26: Legacy GPT model snapshots*, lists davinci-002 and babbage-002 for shutdown that day. On or after that date: whether the section has moved below *Past deprecations*, whether the model pages for davinci-002 and babbage-002 still exist, and whether any model remaining in OpenAI's list is described as not trained with instruction following. The fine-tuned versions follow on 23 October 2026.
- **From 17 September 2026 — DeepSeek-V4.1-Flash-Base.** The model card at huggingface.co/deepseek-ai/DeepSeek-V4.1-Flash reports benchmark results for a base version that, on 17 September, had no public repository. A repository named DeepSeek-V4.1-Flash-Base appearing under the deepseek-ai organisation would restore the practice DeepSeek followed for V4 in April.
- **From 17 September 2026 — a Qwen3.8 base checkpoint.** On that date the newest base language models under huggingface.co/Qwen were the Qwen3.5 sizes of February 2026, and the Qwen3.8-27B card described its weights as the "*post-trained model*". A repository ending in *-Base* for any Qwen3.8 size would show the practice resuming at Qwen.

7. The idea to keep

Supervised fine-tuning is the same objective applied to different text. A base model continues documents; an assistant is the same kind of machine trained further on one kind of document, in which a request is followed by a helpful answer. Instruction tuning extends the habit across many tasks at once. In the two cases where a ratio was published, both from 2021 and 2022, the tuning was under 2% of pre-training; the 2026 model cards of DeepSeek, Qwen, Moonshot and Z.ai describe their post-training without such a ratio, and whether the supervised stage mostly teaches format or something more is argued on 2023 evidence from both directions. A surprising behaviour in an assistant therefore invites a specific question: whether it came from what the model read in pre-training, or from what it was shown afterwards.

Sources

Source	Date	Used for
OpenAI, <i>Aligning Language Models to Follow Instructions</i> (Internet Archive capture of 1 January 2023)	27 January 2022	Paired GPT-3 and InstructGPT outputs; "less than 2%"; "unlocks"
Long Ouyang et al. (OpenAI), <i>Training language models to follow instructions with human feedback</i> , arXiv 2203.02155 v1	4 March 2022	Dataset sizes, labelers, the FLAN and To comparison, Figure 8, limitations
OpenAI, API deprecations page; davinci-002 and babbage-002 model pages	read 17 September 2026	Shutdown dates; "not trained with instruction following"
Hyung Won Chung et al. (Google), <i>Scaling Instruction-Finetuned Language Models</i> , arXiv 2210.11416 (version 5 of 6 December 2022 read)	first posted 20 October 2022	PaLM and Flan-PaLM examples, Figure 9
Andrew M. Dai and Quoc V. Le (Google), <i>Semi-supervised Sequence Learning</i> , arXiv 1511.01432	4 November 2015	Pre-training as a step before supervised training
Alec Radford et al. (OpenAI), <i>Improving Language Understanding by Generative Pre-Training</i>	2018	Fine-tuning per task
Alec Radford et al. (OpenAI), <i>Language Models are Unsupervised Multitask Learners</i>	2019	The "TL;DR:" prompt; Table 4
Jason Wei et al. (Google Research), <i>Finetuned Language Models Are Zero-Shot Learners</i> , arXiv 2109.01652 v1, and ICLR 2022 version	3 September 2021	Definition of instruction tuning; the model-size condition
Swaroop Mishra et al., <i>Natural Instructions: Benchmarking Generalization to New Tasks from Natural Language Instructions</i> , arXiv 2104.08773 v1 (later retitled <i>Cross-Task Generalization via Natural Language Crowdsourcing Instructions</i>)	18 April 2021	Natural Instructions
Victor Sanh et al., <i>Multitask Prompted Training Enables Zero-Shot Task Generalization</i> , arXiv 2110.08207 v1	15 October 2021	To
OpenAI, <i>ChatGPT: Optimizing Language Models for Dialogue</i> (Internet Archive capture of 1 January 2023)	30 November 2022	ChatGPT's supervised first step
Stanford CRFM, <i>Alpaca</i> , and its training code (read 17 September 2026)	13 March 2023	Prompt tokens excluded from the loss
Allen Institute for AI, open-instruct training code (read 17 September 2026)	2026	Non-assistant turns excluded from the loss
Nathan Lambert et al. (AI2), <i>Tulu 3</i> , arXiv 2411.15124 (version 5 of 14 April 2025 read)	first posted 22 November 2024	An open post-training recipe; Table 7
Chunting Zhou et al. (Meta AI and others), <i>LIMA: Less Is More for Alignment</i> , arXiv 2305.11206 v1	18 May 2023	The Superficial Alignment Hypothesis; end-of-turn token
Arnav Gudibande et al. (UC Berkeley), <i>The False Promise of Imitating Proprietary LLMs</i> , arXiv 2305.15717 v1	25 May 2023	Style without factuality

Source	Date	Used for
Bill Yuchen Lin et al. (AI2, University of Washington), <i>The Unlocking Spell on Base LLMs</i> , arXiv 2312.01552 v1	4 December 2023	Token distribution shift; Figure 3
Hugging Face repositories and model cards: DeepSeek, Google, Qwen, Moonshot, Z.ai, Allen Institute for AI, OpenAI	read 17 September 2026	Which releases publish a base checkpoint; GLM-5.3 and Qwen3.5 cards