
Minstral 3



Abstract

We introduce the Minstral 3 series, a family of parameter-efficient dense language models designed for compute and memory constrained applications, available in three model sizes: 3B, 8B, and 14B parameters. For each model size, we release three variants: a pretrained base model for general-purpose use, an instruction finetuned, and a reasoning model for complex problem-solving. In addition, we present our recipe to derive the Minstral 3 models through Cascade Distillation, an iterative pruning and continued training with distillation technique. Each model comes with image understanding capabilities, all under the Apache 2.0 license.

Webpage: <https://mistral.ai/news/mistral-3>

Models: <https://huggingface.co/collections/mistralai/minstral-3>

1 Introduction

In this work, we introduce Minstral 3, a family of dense models trained in a compute- and data-efficient manner through iterative shrinking and distillation from a parent pretrained model. Unlike popular pretrained models such as Qwen3 [Yang et al., 2025] or Llama3 [Dubey et al., 2024] that are trained on 36 trillion and 15 trillion tokens respectively, we are able to produce competitive models trained for between 1 and 3 trillion tokens by leveraging Mistral Small 3.1, a strong 24B-parameter parent model.

Available in three sizes: 3B, 8B, and 14B parameters, all Minstral 3 models are descendants of Mistral Small 3.1¹, obtained via a Cascade Distillation approach. We present three variants for each model size: base, instruct, and reasoning, each with image understanding capabilities and context lengths up to 256k tokens (128k for reasoning models).

A key component of Minstral 3 is our Cascade Distillation training strategy, an iterative pruning and distillation method, which progressively transfers pretrained knowledge from a large parent model down to a family of compact children models. Our recipe allows us to achieve performance that is competitive with models which had a much larger training budget. For example, the Minstral 3 14B Base model closely matches Mistral Small 3.1 Base, while being more than 40% smaller and trained on a much shorter horizon.

After post-training, we achieve competitive results with similarly sized open weight models such as Gemma 3 [Kamath et al., 2025], Qwen 3 [Yang et al., 2025, Bai et al., 2025], and Mistral Small 3.2 2506.

¹<https://mistral.ai/news/mistral-small-3-1>

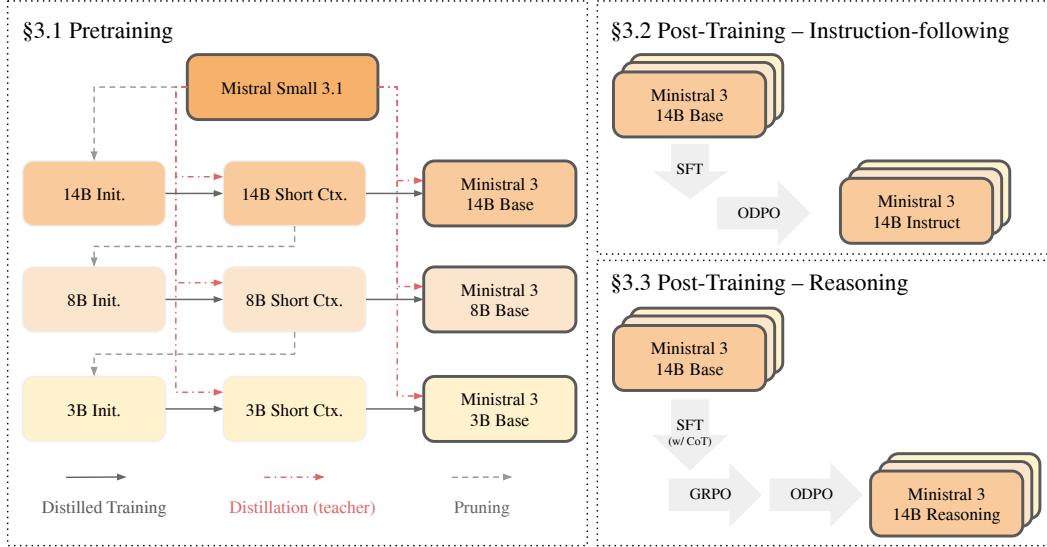


Figure 1: Overview of Minstral 3 training recipe. **Pretraining:** We start from pruning the parent model, Mistral Small 3.1, into the largest child model (14B Init.). Next, we continue pretraining the child model with logit distillation from the parent model as the teacher to obtain the up-trained short context child model (14B Short Ctx.). From 14B Short Ctx., we perform another round of distillation with longer context window (see §3.1 for details) to obtain the final Minstral 3 14B Base model. In parallel, 14B Short Ctx. is pruned to initialize the next child model (8B Init.), from which we repeat the process to derive Minstral 3 8B Base model. We repeat the same process for the 3B version. **Post-training:** Each Base model is then post-trained into the instruction-following and reasoning variants. For instruction-following, our post-training recipe includes supervised fine-tuning (SFT) and Online Direct Preference Optimization (ODPO). For reasoning, the process involved supervised fine-tuning with chain-of-thought data (SFT w/ CoT), Group Relative Policy Optimization (GRPO; Shao et al. [2024]), and ODPO.

The main contributions can be summarized as follows:

- We introduce Minstral 3, a family of 9 dense language models - a pretrained, an instruction finetuned, and a reasoning model, each at the 14B, 8B, and 3B parameter scales. All Minstral 3 models (3 sizes \times 3 variants) are open-weight under the Apache 2.0 license.
- We present a compute-efficient pretraining recipe, Cascade Distillation, with which these models have been pretrained at a fraction of the cost it would take to pretrain from scratch.
- We independently confirm findings from prior work that (a) there exists a "capacity gap" where a stronger teacher does not yield a stronger student model for pretraining, but post-training continues to benefit from a stronger teachers (b) distilling from a post-trained as opposed to a pretrained teacher when pretraining the student model results in better benchmark scores (c) distilling from a human preference optimized teacher is better than one that has only been post-trained with SFT.

2 Model Architecture

Table 1: Architectural specifications and hyperparameters for the Minstral 3 family. All models use a vocabulary size of 131K tokens.

	Layers	Latent dim.	Q/KV heads	FFN dim.	Tied Embeddings	Context Length
Minstral 3 14B	40	5120	32 / 8	16384	✗	256k
Minstral 3 8B	34	4096	32 / 8	14336	✗	256k
Minstral 3 3B	26	3072	32 / 8	9216	✓	256k

The Minstral 3 family is based on the decoder-only transformer architecture [Vaswani et al., 2017]. All models share a common architectural foundation with size-specific scaling. As shown in Table 1, the family consists of three sizes: 3B, 8B, and 14B parameters, with 26, 34, and 40 layers

respectively. Other architectural choices include Grouped Query Attention [Ainslie et al., 2023] with 32 query heads and 8 key-value heads, RoPE [Su et al., 2021] positional embeddings, SwiGLU activation [Shazeer, 2020], and RMSNorm [Zhang and Sennrich, 2019]. For long-context extension, we use YaRN [Peng et al., 2023] and position-based softmax temperature scaling in the attention layer [Nakanishi, 2025, MetaAI, 2025]. The 3B model uses tied input-output embeddings to avoid embedding parameters dominating the overall parameter count. All models use a vocabulary of 131K tokens and support context lengths up to 256K tokens.

Vision encoder. All Minstral 3 models use a 410M parameter ViT as a vision encoder for image understanding that is copied from Mistral Small 3.1 Base and kept frozen, with the same architecture described in Pixtral [Agrawal et al., 2024]. We discard the pretrained projection layer from the ViT to language model’s space and train a new projection for every model.

3 Training Recipe

Figure 1 illustrates the training pipeline of the Minstral 3 models, consisting of a pretraining followed by two distinct post-training phases to produce instruction finetuned and reasoning variants.

3.1 Pretraining

Algorithm 1 Cascade Distillation.

```

1 model = MS3 # Mistral Small 3.1
2 for model_size in [14B, 8B, 3B]:
3     # pruning (see Algo. 2)
4     model = prune(model, model_size)
5     # short context distillation
6     model = model.train(
7         data=short_data,
8         teacher_model=MS3,
9     )
10    # long context distillation
11    final_model = model.train(
12        data=long_data,
13        teacher_model=MS3,
14    )
15    yield (model_size, final_model)
16
17
18
19

```

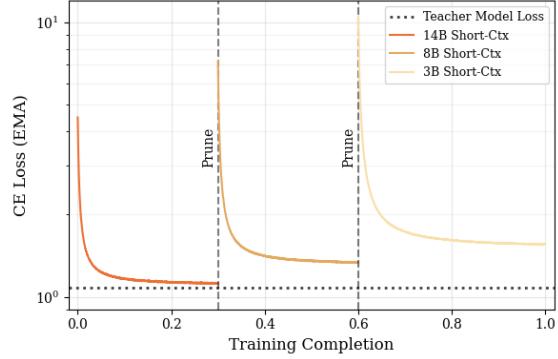


Figure 2: Illustration of Cascade Distillation.

Cascade Distillation. Pretraining of the Minstral 3 models starts from the Mistral Small 3.1 Base (MS3.1) model. We use *Cascade Distillation*, an iterative approach to prune and distill MS3.1 into the smaller successors. Cascade Distillation is a compute-efficient process for pretraining children models of decreasing target sizes, given a pre-trained larger parent model. As summarized in Algorithm 1, it relies on an iterative “prune-distill-repeat” approach:

1. Prune: initialize the weights of a child model via pruning a larger pre-trained model.
2. Distill: up-train the freshly pruned model via distillation from the teacher model’s logits.
3. Repeat: apply this strategy repeatedly to shrink the child model into something even smaller.

Model pruning at each stage follows a similar approach to Minitron and Wanda [Sun et al., 2023, Sreenivas et al., 2024, Muralidharan et al., 2024] with the distillation teacher being Mistral Small 3.1 for all variants. Details of pruning and distillation are provided in the following paragraphs.

Compared to training each small model from scratch, Cascade Distillation produces a model that is significantly more FLOP efficient. It is also worth noting that the end-to-end process can be viewed as a form of continual pretraining of the parent model with weight pruning. As illustrated in Figure 2, data repetition is avoided throughout the process as Cascade Distillation goes through the data mix in a single run with pruning en route.

Pruning. Similar to Minitron, our pruning strategies are designed to preserve the most critical components of the original model (over a validation dataset) while reducing its size. We employ following key pruning techniques:

- **Layer Pruning:** Unlike Sreenivas et al. [2024], which relies on counterfactual downstream perplexities from removing individual layers, we find that the ratio of input to output activation norms provides a simpler yet strong proxy for layer importance.
- **Hidden Dimension Pruning:** Apply Principal Component Analysis (PCA) to concatenated activations from attention normalization and feed-forward normalization layers across all layers. This yields a single rotation matrix consistent across the entire network that projects the model to a lower-dimensional space while maximizing explained variance.
- **Feedforward Dimension Pruning:** For MLPs with gated-linear activation functions such as SwiGLU [Shazeer, 2020], expressed as $W_2(SiLU(W_1x) * W_3x)$ given a very large batch x , we prune dimension of the matrices W_1, W_2, W_3 . To determine the columns of W_1, W_3 to keep, we compute the importance score defined as the averaged absolute value of each dim of the expression above. We then keep only the corresponding rows of W_2 with the indices yielded above.

Algorithm 2 provides more detail on our pruning strategy:

Algorithm 2 Pruning stage of Cascade Distillation. It takes as input a pre-trained model and target size configuration to prune to. We use `input_x` and `output_x` to refer to activations from a large calibration batch.

```

1 def prune(model, target_size):
2     target_n_layers, target_dim, target_ffn_dim = get_config(target_size)
3
4     # layer pruning
5     scores = []
6     for layer in model.layers:
7         input_norm = layer.input_x.norm(dim=-1)
8         output_norm = layer.output_x.norm(dim=-1)
9         scores.append(
10             (output_norm / input_norm).mean()
11         )
12
13     layers_to_keep = topk(scores, k=target_n_layers)
14     model = remove_layers(model, layers_to_keep)
15
16     # hidden dimension pruning
17     norm_inputs = []
18     for layer in model.layers:
19         norm_inputs.extend([
20             layer.attn_norm.input_x,
21             layer.ffn_norm.input_x,
22         ])
23
24     rotation = PCA(norm_inputs, n_components=n_dims)
25     model = apply_rotation(model, rotation, target_dim)
26
27     # feedforward pruning
28     for layer in model.layers:
29         importance = abs(
30             silu(layer.ffn.w1.output_x) * layer.ffn.w3.output_x
31         ).mean(dim=(0,1))
32         dims_to_keep = topk(importance, k=target_ffn_dim)
33         layer.ffn = prune_hidden_dims(layer.ffn, dims_to_keep)
34
35
36
37 return model

```

Distillation. After weight initialization, each child model is trained on a mixture of text-only and interleaved text with image data with logit distillation from a teacher model. We find that training with just the forward KL distillation objective outperforms tuning the coefficients of an objective that weights the distillation objective and the next token prediction objective differently. For all stages and model sizes, we use the parent model as the teacher model (more details in §5.1).

The pretraining phase consists of a two-stages:

- (1) **Short context stage** with a context window of length 16,384. The output of this phase is the input to the pruning phase of the next child model.

- (2) **Long context stage** to extend the context window from 16,384 to 262,144 using YaRN [Peng et al., 2023] and position-based temperature scaling [Nakanishi, 2025, MetaAI, 2025].

3.2 Post-Training: Minstral Instruct

To impart instruction-following capabilities [Ouyang et al., 2022], pretrained models are fine-tuned using a curated dataset comprising high-quality multimodal and text-only instruction-following data. The fine-tuning phase also consists of two stages: Supervised Fine-Tuning (SFT) and Online Direct Preference Optimization (ODPO).

3.2.1 Supervised Fine-tuning

We run SFT with fp8 quantization, using a logit distillation loss from a strong teacher. Unlike pretraining, each model is distilled from Mistral Medium 3 model (more details in §5.1). Similar to the pretraining phase, the vision encoder remains frozen while the adapter is trainable.

3.2.2 Online Direct Preference Optimization stage

Direct Preference Optimization (DPO) [Rafailov et al., 2023] offers a lightweight framework for human preference optimization by learning directly from offline pairwise preferences. For the Minstral 3 models, we adopt its online variant, Online Direct Preference Optimization (ODPO) [Guo et al., 2024] where, for each example, we sample two candidate responses from the current policy with temperature $T=0.7$, and use a text-based reward model to rank the responses.

This method relies on a Pairwise Reward Model (PWRM) to dynamically rank candidate responses. The PWRM is trained via supervised fine-tuning (SFT) on structured pairwise data: given a conversation history and two candidate responses, it predicts which response is preferred. In addition, we refine the classic DPO loss by incorporating the binomial probabilistic output of the PWRM, replacing hard winner/loser labels with a two-sided loss that weights each response by its probability of being preferred. We make two additional changes to stabilize the learning process: (1) we adjust the PWRM temperature to calibrate the win / loss probabilities; and (2) we employ a β -rescaling technique, allowing for a more beta-invariant rescaling of dpo loss.

In practice, the online variant is particularly important for mitigating model-induced artifacts, such as infinite generations. This is also facilitated by some heuristics, such as automatically treating any response that exhibits an infinite loop during sample as “loser,” preventing such behavior from being reinforced. Finally, we enable tool execution during generation, which improves the model’s tool-use performance.

In summary, we found that using online preference optimization improves alignment with human preferences significantly over both the SFT and offline variants. We release the models resulting from this phase as Minstral 3-14B/8B/3B Instruct.

3.3 Post-Training: Minstral Reasoning

Post-training for reasoning models begins from the pre-trained checkpoint as opposed to the ODPO variant. We train the model for inference-time scaling using a three-stage pipeline composed of SFT, GRPO and ODPO, using the long-context pretrained checkpoint as the starting point. Models released after this reasoning-oriented fine-tuning stage are referred to as Minstral 3 14B/8B/3B Reasoning.

3.3.1 Reasoning Supervised Fine-Tuning

In this stage, the model is finetuned on a mixture of short and long CoT samples. The former is derived from our general SFT data mixture whereas the latter consists of reasoning traces which have been prefixed with a reasoning specific system prompt.

The reasoning traces come from a diverse set of domains including mathematics, coding, general dialogue, instruction following, multilingual tasks, tool use, and visual reasoning. We apply lightweight filtering to remove examples that are poorly formatted, contain excessive repetition, or have undesirable language switching, ensuring that the model is exposed to clean and well-structured chains of thought.

3B SFT: For the 3B model, vanilla SFT led to a brittle, overly verbose model with lots of repetition and infinite generations in its output. To mitigate this, we did logit distillation with Magistral Small 1.2 as teacher. This helped reduce verbosity and stabilized subsequent RL training.

3.3.2 Reinforcement Learning

We perform GRPO [DeepSeek-AI et al., 2025] on top of the SFT checkpoint to refine the model’s thinking and improve the performance further on reasoning tasks. The training is conducted in two stages:

STEM RL: In the first stage, we train the model on math, code and visual-reasoning tasks. We collect question-answer pairs from a diverse set of open and proprietary sources. The samples are filtered and cleaned using a rigorous multi-step pipeline (detailed in Rastogi et al. [2025]) to remove invalid, incomplete and very easy/hard problems.

General RL: In the second stage, we broaden the scope beyond STEM problems. We generate atomic grading rubrics for a diverse set of prompts including general chat, instruction-following, and open-ended reasoning tasks. During GRPO, an LLM judge evaluates each model rollout against these rubrics (e.g., faithfulness to the prompt, response quality) and the final reward is set to the fraction of satisfied heuristics. This stage improves the instruction following and general chat capabilities of the model while maintaining, and sometimes even improving, the performance on the STEM benchmarks.

For both stages, we follow the GRPO training recipe from Rastogi et al. [2025]. The maximum generation length is increased from 32K to 80K, since we observed a non-trivial proportion of truncated generations during RL. Allowing longer outputs allowed the model to finish its reasoning for the most challenging problems, resulting in additional performance gains.

3.3.3 Online Direct Preference Optimization

Finally, we apply ODPO as a post-RL alignment stage to better align with user preferences and polish the model’s conversational and instructional behavior. The overall procedure follows the same setup as used for our non-reasoning instruct models, with one modification – The thinking chunks are stripped from the model’s generations before sending them to the reward model for scoring. Some additional experimental details are discussed in Section 5.3.

4 Results

In this section, we report the results of Minstral 3 models on a variety of benchmarks. We also compare Minstral 3 to other open-weight models on the same scale, namely the Qwen 3 family [Yang et al., 2025, Bai et al., 2025] and the Gemma 3 family [Kamath et al., 2025]. For external models, we re-run all benchmarks with our own evaluation pipeline for fair comparison.

We evaluated on the following benchmarks: **General:** MMLU [Hendrycks et al., 2020], MMLU-Redux [Perez et al., 2024], ARC-Challenge [Clark et al., 2018], RACE High [Lai et al., 2017], TriviaQA [Joshi et al., 2017], NaturalQS [Kwiatkowski et al., 2019], and AGIEval [Zhong et al., 2023]. **Math & Code:** MATH [Hendrycks et al., 2021], GPQA Diamond [Rein et al., 2024], and MBPP [Austin et al., 2021]. **Multimodal:** MMMU [Yue et al., 2024] and MathVista [Lu et al., 2024]. **Post-training:** Arena Hard [Li et al., 2024], WildBench [Lin et al., 2024], MM MTBench², AIME 2024/2025, HMMT 2025, PhyBench [Liu et al., 2025], and LiveCodeBench [Jain et al., 2024].

4.1 Pretraining Results

In Table 2, we compare Minstral 3 Base models against other open-weight models of similar size from the Gemma 3 family and the Qwen 3 family.

At the 14B scale, Minstral 3 demonstrates strong performance, outperforming Qwen 3 14B on TriviaQA and MATH, while being competitive on other benchmarks. Our 14B model is also significantly better than Gemma 12B across all benchmarks. At the 8B scale, we observe a similar trend. It is also worth pointing out that Minstral 3 8B outperforms the larger Gemma 12B in most of the evaluations (except TrivaiQA), highlighting the strong parameter efficiency of Minstral 3 models.

²<https://huggingface.co/datasets/mistralai/MM-MT-Bench>

Table 2: Comparing Minstral 3 Base models against the Gemma 3 base models and the Qwen 3 base models on pretraining benchmarks. All the results are reported after running the evaluations using our internal harness with identical configuration.

Model	MMLU-Redux (5-shot)	TriviaQA (5-shot)	MATH (CoT 2-Shot)	AGIEval (5-shot)	Multilingual MMLU (5-Shot)
Qwen 3 14B	83.7	70.3	62.0	66.1	75.4
Minstral 3 14B	82.0	74.9	67.6	64.8	74.2
Gemma 3 12B	76.6	78.8	48.7	58.7	69.0
Qwen 3 8B	79.4	63.9	57.6	59.6	70.0
Minstral 3 8B	79.3	68.1	62.6	59.1	70.6
Gemma 3 4B	62.6	64.0	29.4	43.0	51.6
Qwen 3 4B	75.9	53.0	40.5	57.0	67.7
Minstral 3 3B	73.5	59.2	60.1	51.1	65.2

Table 3: Evaluation results of the Minstral 3 Base family compared to the teacher model Mistral Small 3.1 24B across general reasoning, math & code, multilingual, and multimodal benchmarks. Performance scales smoothly with model size, yet the pruned Minstral 3 variants retain a large fraction of the teacher’s capability despite substantial parameter reductions.

Evaluation	Mistral Small 24B	Minstral 3 14B	Minstral 3 8B	Minstral 3 3B
<i>General</i>				
MMLU (5-shot)	81.0	79.4	76.1	70.7
MMLU-Redux (5-shot)	82.7	82.0	79.3	73.5
ARC-Challenge	91.6	89.9	88.0	85.5
RACE High	52.1	52.3	49.7	49.3
TriviaQA (5-shot)	79.3	74.9	68.1	59.2
NaturalQS (5-shot)	34.4	29.9	25.8	21.9
<i>Math & Code</i>				
MATH (CoT 2-Shot)	55.8	67.6	62.6	60.1
GPQA Diamond (0-shot)	36.9	39.9	39.9	33.8
MBPP (3-shot Pass@1)	71.6	71.6	70.0	63.0
<i>Multilingual MMLU</i>				
European avg. [†] (5-shot)	78.8	76.9	73.4	68.4
Chinese (5-shot)	75.7	75.1	71.3	64.1
Japanese (5-shot)	76.7	75.9	72.2	65.7
Korean (5-shot)	59.3	59.0	55.3	48.9
<i>Multimodal</i>				
MMMU (2-shot)	59.1	59.9	55.1	52.4
MathVista	51.3	43.6	35.7	23.3

[†] Averaged over German, Spanish, French, Italian, and Portuguese.

At the 3B scale, the same overall trend persists, but performance gaps between models become more pronounced. Additional pretraining evaluation results for Minstral 3 Base models along with the teacher model are provided in Table 3.

4.2 Post-training Results

In Table 4, we compare Minstral 3 Instruct models against Instruct models from the Gemma 3 family and the Qwen 3 family. For Qwen 3, we report the results for the latest vision enabled instruct variants (Qwen3-VL).

Table 4: Performance comparison of Minstral 3 instruct models against instruction-tuned baselines from the Qwen 3 and Gemma 3 families. Models are grouped by size to facilitate like-for-like comparisons.

Model	Arena Hard	WildBench	MATH (maj@1)	MM MTBench
Qwen3 14B (Non-Thinking)	42.7	65.1	87.00	N/A
Minstral 3 14B	55.1	68.5	90.40	84.90
Gemma3-12B-Instruct	43.6	63.2	85.40	67.00
Qwen3-VL-8B-Instruct	52.8	66.3	94.60	80.00
Minstral 3 8B	50.9	66.8	87.60	80.80
Gemma3-4B-Instruct	31.8	49.1	75.90	52.30
Qwen3-VL-4B-Instruct	43.8	56.8	90.00	80.08
Minstral 3 3B	30.5	56.8	83.00	78.30
Qwen3-VL-2B-Instruct	16.3	42.2	78.60	63.60

Table 5: Comparison of Minstral 3 reasoning models with size-matched Qwen 3 reasoning counterparts on mathematics, science, and code benchmarks.

Benchmark	Qwen 3 14B	Minstral 3 14B	Qwen3-VL 8B	Minstral 3 8B	Qwen3-VL 4B	Minstral 3 3B
AIME 2024	83.7	89.8	86.0	86.0	72.9	77.5
AIME 2025	73.7	85.0	79.8	78.7	69.7	72.1
HMMT 2025	55.8	67.5	57.5	55.8	50.8	51.7
GPQA Diamond	66.3	71.2	67.1	66.8	60.1	53.4
PhyBench	22.0	26.0	22.0	20.0	9.0	15.0
LiveCodeBench v6	59.3	64.6	58.0	61.6	51.3	54.8

In Table 5, we compare Minstral 3 Reasoning models against reasoning models from the Qwen 3 family. To ensure a fair comparison, all models are evaluated using the same evaluation pipeline. To reduce variance, we report pass@16 except LiveCodeBench which is evaluated using pass@5.

5 Discussions

5.1 Choice of Teacher Model for Distillation

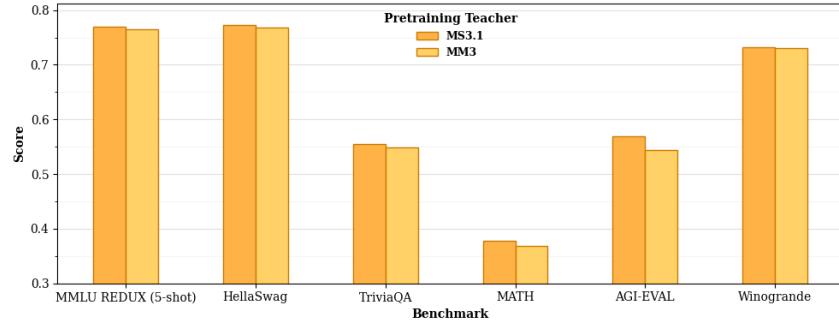


Figure 3: Minstral 3 14B pretraining ablations comparing distillation from Mistral Small 3.1 and Mistral Medium 3 teachers. Despite Mistral Medium 3 being larger and more capable, distillation from Mistral Small 3.1 consistently yields stronger downstream performance across different benchmarks.

In selecting an appropriate teacher model for the distillation process, we identified several noteworthy observations that meaningfully influenced our design choices:

Stronger teacher does not lead to better results: For pretraining, distilling from Mistral Small 3.1 outperformed distillation from the much stronger Mistral Medium 3³ even in a non FLOP-matched setup, similar to observations in Busbridge et al. [2025] (Figure 3). However, during post-training, Minstral 3 models benefit from distillation from the more capable Mistral Medium 3.1.

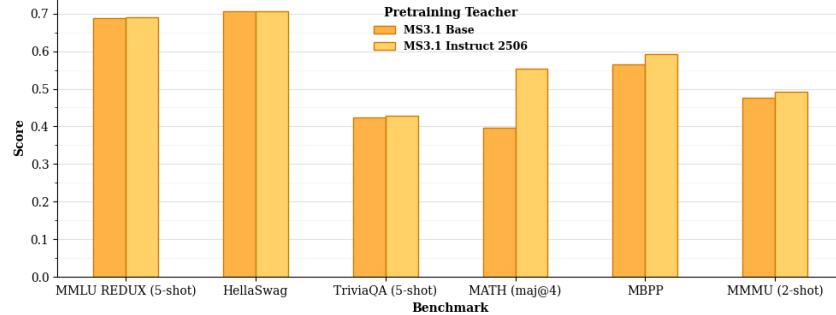


Figure 4: Minstral 3 3B pretraining ablations comparing distillation from base and post-trained (instruct/reasoning) variants of Mistral Small 3.1. The instruct teacher yields stronger performance on STEM benchmarks, while achieving comparable results on knowledge and multimodal evaluations..

The choice of teacher version (base / instruct) matters: In line with Goyal et al. [2025], we find that distilling from a post-trained teacher as opposed to a pre-trained one during the pre-training stage results in a stronger model (Figure 4). In particular, this had a strong impact on maths (MATH) and code capabilities, a small but consistent impact on multimodal evaluations (e.g. MMU), and a negligible impact on knowledge metrics (MMLU / Trivia-QA).

Human Preference tuned models are better teachers: Post We use two internal versions of Mistral Medium 3 to answer the question - is it better to distill from an SFT or a preference tuned checkpoint during SFT? We find that distilling from the preference tuned checkpoint is always substantially better. These gains persist even after the student model undergoes its own preference tuning phase.

5.2 Model Verbosity.



Figure 5: Verbosity (in terms of number of output tokens) v.s. accuracy on GPQA Diamond with Minstral 3 instruction-following and reasoning.

Our post-training of Minstral 3 Instruct differs from Qwen 3 in that it does not do "Reasoning RL" before the "General RL" stage (see Fig. 1 of Yang et al. [2025]) this likely results in different model verbosity between the two models as illustrated in Figure 5.

³<https://docs.mistral.ai/models/mistral-medium-3-1-25-08>

In an experiment to try and get the Minstral 3 Instruct models to produce longer chains of thought, we investigated incorporating varying proportions of long chain-of-thought (CoT) reasoning traces, paired with carefully curated system prompts, into the SFT training data. Increasing the fraction of such Long CoT data improved the performance on STEM benchmarks; however, it also leads to excessive reflection, internal monologues and backtracking behavior (as shown below), which is undesirable and unnatural for a general-purpose chat model.

Reasoning Behavior in Instruct Model

Okay, let's try to solve this problem step by step. The question is: Find the sum of all integer bases $b > 9$ for which 17_b is a divisor of 97_b . First, I need to recall how to convert numbers from base b to base 10. A number $d_nd_{n-1}\dots d_1d_0$ in base b is equal to $d_nb^n + d_{n-1}b^{n-1} + \dots + d_1b + d_0$ in base 10.

...

Wait, the problem states $b > 9$, so we need to check the divisors of 56 that are greater than 9 (since $b = d - 7$)

...

Wait, but let's check another way. Instead of using the congruence, we can set up the equation $9b + 7 = k(b + 7)$ for some integer $k \geq 1$

...

I think that's it. The key was to convert the numbers to base 10, set up the divisibility condition, and recognize that the base must be greater than 9, so we only consider divisors of 56 that make $b > 9$.

...

Thus, the valid bases are 21 and 49. The sum of these bases is:

$$21 + 49 = 70$$

Therefore, the answer is 70.

5.3 ODPO for Minstral 3 Reasoning.

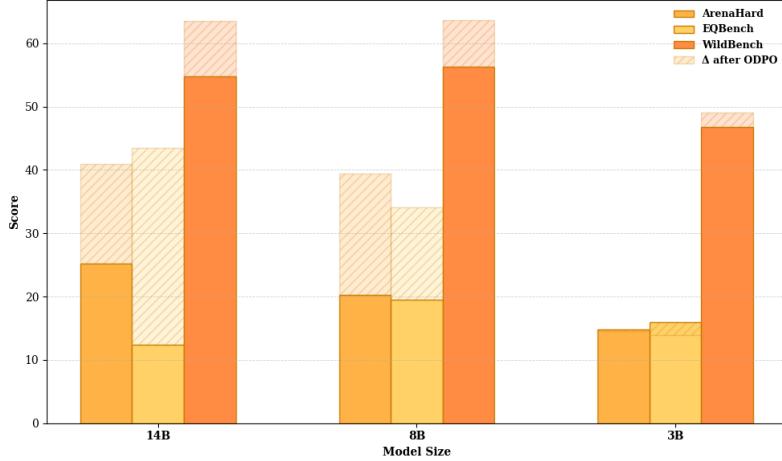


Figure 6: Impact of ODPO on chat benchmarks for Minstral 3 reasoning models, applied on top of GRPO-trained checkpoints. ODPO delivers substantial gains across all benchmarks for the 14B and 8B variants.

Reasoning models, while being better at solving challenging problems, often lag in general conversational quality, a pattern we also observed with the Minstral 3 reasoning variants. To address this, we performed ODPO training on top of the RL-trained checkpoints. As shown in Figure 6, this significantly improved the 14B and 8B models on alignment benchmarks. The 3B model however, did not demonstrate significant improvements on public benchmarks after this stage⁴. The model nevertheless performed better in our internal human evaluations and so we selected the ODPO checkpoint as the release candidate.

⁴We also found the 3B base more sensitive than 14B and 8B to hyper-parameter choice in fine-tuning

6 Conclusion

We introduced Minstral 3, a family of efficient dense language models designed for resource-constrained environments. Through iterative distillation from larger teacher models (Mistral Small 3.1 and Medium 3), we created three model sizes (14B, 8B, 3B) each available in base, instruction-following, and reasoning-enhanced variants. All models support vision capabilities and handle contexts up to 256K tokens. Collectively, Minstral 3 models highlight Mistral’s continued commitment to supporting and advancing open-source initiatives. We hope they will provide value to the community and contribute to a stronger, more vibrant open-source ecosystem.

Core contributors

Alexander H. Liu, Kartik Khandelwal, Sandeep Subramanian, Victor Jouault

Contributors

Abhinav Rastogi, Adrien Sadé, Alan Jeffares, Albert Jiang, Alexandre Cahill, Alexandre Gavaudan, Alexandre Sablayrolles, Amélie Héliou, Amos You, Andy Ehrenberg, Andy Lo, Anton Eliseev, Antonia Calvi, Avinash Sooriyarachchi, Baptiste Bout, Baptiste Rozière, Baudouin De Monicault, Clémence Lanfranchi, Corentin Barreau, Cyprien Courtot, Daniele Grattarola, Darius Dabert, Diego de las Casas, Elliot Chane-Sane, Faruk Ahmed, Gabrielle Berrada, Gaëtan Ecrepont, Gauthier Guinet, Georgii Novikov, Guillaume Kunsch, Guillaume Lample, Guillaume Martin, Gunshi Gupta, Jan Ludziejewski, Jason Rute, Joachim Studnia, Jonas Amar, Joséphine Delas, Josselin Somerville Roberts, Karmesh Yadav, Khyathi Chandu, Kush Jain, Laurence Aitchison, Laurent Fainsin, Léonard Blier, Lingxiao Zhao, Louis Martin, Lucile Saulnier, Luyu Gao, Maarten Buyl, Margaret Jennings, Marie Pellat, Mark Prins, Mathieu Poirée, Mathilde Guillaumin, Matthieu Dinot, Matthieu Futeral, Maxime Darrin, Maximilian Augustin, Mia Chiquier, Michel Schimpf, Nathan Grinsztajn, Neha Gupta, Nikhil Raguraman, Olivier Bousquet, Olivier Duchenne, Patricia Wang, Patrick von Platen, Paul Jacob, Paul Wambergue, Paula Kurylowicz, Pavankumar Reddy Muddireddy, Philomène Chagniot, Pierre Stock, Pravesh Agrawal, Quentin Torroba, Romain Sauvestre, Roman Soletskyi, Rupert Menneer, Sagar Vaze, Samuel Barry, Sanchit Gandhi, Siddhant Waghjale, Siddharth Gandhi, Soham Ghosh, Srijan Mishra, Sumukh Aithal, Szymon Antoniak, Teven Le Scao, Théo Cachet, Theo Simon Sorg, Thibaut Lavril, Thiziri Nait Saada, Thomas Chabal, Thomas Foubert, Thomas Robert, Thomas Wang, Tim Lawson, Tom Bewley, Tom Bewley, Tom Edwards, Umar Jamil, Umberto Tomasini, Valeria Nemychnikova, Van Phung, Vincent Maladière, Virgile Richard, Wassim Bouaziz, Wen-Ding Li, William Marshall, Xinghui Li, Xinyu Yang, Yassine El Ouahidi, Yihan Wang, Yunhao Tang, Zaccharie Ramzi

References

- Pravesh Agrawal, Szymon Antoniak, Emma Bou Hanna, Baptiste Bout, Devendra Chaplot, Jessica Chudnovsky, Diogo Costa, Baudouin De Monicault, Saurabh Garg, Theophile Gervet, et al. Pixtral 12b. *arXiv preprint arXiv:2410.07073*, 2024.
- Joshua Ainslie, James Lee-Thorp, Michiel de Jong, Yury Zemlyanskiy, Federico Lebrón, and Sumit Sanghi. Gqa: Training generalized multi-query transformer models from multi-head checkpoints. *arXiv preprint arXiv:2305.13245*, 2023.
- Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, et al. Program synthesis with large language models. *arXiv preprint arXiv:2108.07732*, 2021.
- Shuai Bai et al. Qwen3-vl technical report, 2025. URL <https://arxiv.org/abs/2511.21631>.
- Dan Busbridge, Amitis Shidani, Floris Weers, Jason Ramapuram, Eta Littwin, and Russ Webb. Distillation scaling laws. *arXiv preprint arXiv:2502.08606*, 2025.
- Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. Think you have solved question answering? try arc, the ai2 reasoning challenge. *arXiv preprint arXiv:1803.05457*, 2018.
- DeepSeek-AI et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. URL <https://arxiv.org/abs/2501.12948>.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv e-prints*, pages arXiv–2407, 2024.
- Sachin Goyal, David Lopez-Paz, and Kartik Ahuja. Distilled pretraining: A modern lens of data, in-context learning and test-time scaling, 2025. URL <https://arxiv.org/abs/2509.01649>.
- Shangmin Guo, Biao Zhang, Tianlin Liu, Tianqi Liu, Misha Khalman, Felipe Llinares, Alexandre Rame, Thomas Mesnard, Yao Zhao, Bilal Piot, Johan Ferret, and Mathieu Blondel. Direct language model alignment from online ai feedback, 2024. URL <https://arxiv.org/abs/2402.04792>.
- Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*, 2020.
- Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*, 2021.
- Naman Jain, King Han, Alex Gu, Wen-Ding Li, Fanjia Yan, Tianjun Zhang, Sida Wang, Armando Solar-Lezama, Koushik Sen, and Ion Stoica. Livecodebench: Holistic and contamination free evaluation of large language models for code. *arXiv preprint arXiv:2403.07974*, 2024.
- Mandar Joshi, Eunsol Choi, Daniel S Weld, and Luke Zettlemoyer. Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension. *arXiv preprint arXiv:1705.03551*, 2017.
- Aishwarya Kamath, Johan Ferret, Shreya Pathak, et al. Gemma 3 technical report, 2025. URL <https://arxiv.org/abs/2503.19786>.
- Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, et al. Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7:453–466, 2019.
- Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. Race: Large-scale reading comprehension dataset from examinations. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 785–794, 2017.

Tianle Li, Wei-Lin Chiang, Evan Frick, Lisa Dunlap, Tianhao Wu, Banghua Zhu, Joseph E Gonzalez, and Ion Stoica. From crowdsourced data to high-quality benchmarks: Arena-hard and benchbuilder pipeline. *arXiv preprint arXiv:2406.11939*, 2024.

Bill Yuchen Lin, Yuntian Deng, Khyathi Chandu, Faeze Brahman, Abhilasha Ravichander, Valentina Pyatkin, Nouha Dziri, Ronan Le Bras, and Yejin Choi. Wildbench: Benchmarking llms with challenging tasks from real users in the wild. *arXiv preprint arXiv:2406.04770*, 2024.

Zihan Liu, Zijian Wang, Yue Zhang, Jianing Wang, Jian Tang, Xiang He, and Xiangyu Zhang. Phybench: Holistic evaluation of physical perception and reasoning in large language models. *arXiv preprint arXiv:2504.16074*, 2025.

Pan Lu, Hritik Bansal, Tony Xia, Jiacheng Liu, Chunyuan Li, Hannaneh Hajishirzi, Hao Cheng, Kai-Wei Chang, Michel Galley, and Jianfeng Gao. Mathvista: Evaluating mathematical reasoning of foundation models in visual contexts. In *International Conference on Learning Representations (ICLR)*, 2024.

MetaAI. The llama 4 herd: The beginning of a new era of natively multimodal ai innovation. <https://ai.meta.com/blog/llama-4-multimodal-intelligence/>, 2025.

Saurav Muralidharan, Sharath Turuvekere Sreenivas, Raviraj Joshi, Marcin Chochowski, Mostofa Patwary, Mohammad Shoeybi, Bryan Catanzaro, Jan Kautz, and Pavlo Molchanov. Compact language models via pruning and knowledge distillation. *arXiv preprint arXiv:2407.14679*, 2024.

Ken M Nakanishi. Scalable-softmax is superior for attention. *arXiv preprint arXiv:2501.19399*, 2025.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

Samuel J. Paech. Eq-bench: An emotional intelligence benchmark for large language models, 2023.

Bowen Peng, Jeffrey Quesnelle, Honglu Fan, and Enrico Shippole. Yarn: Efficient context window extension of large language models. *arXiv preprint arXiv:2309.00071*, 2023.

Aryo Perez, Tomasz Stanislawek, Andrzej Pohl, Kamil Dwojak, Dawid Jurkiewicz, Piotr Kobus, and Tomasz Trzcinski. Are we done with mmlu? *arXiv preprint arXiv:2406.04127*, 2024.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *arXiv preprint arXiv:2305.18290*, 2023.

Abhinav Rastogi, Albert Q. Jiang, Andy Lo, Gabrielle Berrada, Guillaume Lample, et al. Magistral. *arXiv preprint arXiv:2506.10910*, 2025.

David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R Bowman. Gpqa: A graduate-level google-proof q&a benchmark. In *First Conference on Language Modeling*, 2024.

Keisuke Sakaguchi, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi. Winogrande: An adversarial winograd schema challenge at scale. *Communications of the ACM*, 64(9):99–106, 2021.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Huawei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024. URL <https://arxiv.org/abs/2402.03300>.

Noam Shazeer. Glu variants improve transformer. *arXiv preprint arXiv:2002.05202*, 2020.

Sharath Turuvekere Sreenivas, Saurav Muralidharan, Raviraj Joshi, Marcin Chochowski, Ameya Sunil Mahabaleshwarkar, Gerald Shen, Jiaqi Zeng, Zijia Chen, Yoshi Suhara, Shizhe Diao, Chenhan Yu, Wei-Chun Chen, Hayley Ross, Oluwatobi Olabiyi, Ashwath Aithal, Oleksii Kuchaiev, Daniel Korzekwa, Pavlo Molchanov, Mostofa Patwary, Mohammad Shoeybi, Jan Kautz, and Bryan Catanzaro. Llm pruning and distillation in practice: The minitron approach. *arXiv preprint arXiv:2408.11796*, 2024.

Jianlin Su, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *arXiv preprint arXiv:2104.09864*, 2021.

Mingjie Sun, Zhuang Liu, Anna Bair, and J Zico Kolter. A simple and effective pruning approach for large language models. *arXiv preprint arXiv:2306.11695*, 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

An Yang et al. Qwen3 technical report, 2025. URL <https://arxiv.org/abs/2505.09388>.

Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, and Wenhui Chen. Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi. In *Proceedings of CVPR*, 2024.

Rowan Zellers, Ari Holtzman, Yonatan Bisk, Ali Farhadi, and Yejin Choi. Hellaswag: Can a machine really finish your sentence? *arXiv preprint arXiv:1905.07830*, 2019.

Biao Zhang and Rico Sennrich. Root mean square layer normalization. *Advances in Neural Information Processing Systems*, 32, 2019.

Wanjun Zhong, Ruixiang Cui, Yiduo Guo, Yaobo Liang, Shuai Lu, Yanlin Wang, Amin Saied, Weizhu Chen, and Nan Duan. Agieval: A human-centric benchmark for evaluating foundation models. *arXiv preprint arXiv:2304.06364*, 2023.