TRL 1.19 explained: multi-teacher distillation, faster packing and fine-tuning fixes

26 November 2026 · 4 min read · blogs.iictp.org/trl-1-19-multi-teacher-distillation-packing-explained

Short answer: TRL 1.19 is a release of Hugging Face's open-source library for post-training language models, published on 25 November 2026, that adds distillation from several teacher models, distillation against agent environments, faster packed supervised fine-tuning and a fix for fine-tuned models that failed to stop. It also adds support for vLLM 0.34 and raises minimum dependency versions. The speed figures are the maintainers' own measurements.

TL;DR The short version

- TRL 1.19, Hugging Face's post-training library released on 25 November 2026, adds distillation from several teacher models and against agent environments.
- It also speeds up packed supervised fine-tuning, fixes fine-tuned models that failed to stop, and supports vLLM 0.34 with higher minimum dependencies.
- Speed figures are the maintainers' own, so pin versions and rerun your training checks after upgrading.

Read online, listen to it, or change the text size: https://blogs.iictp.org/trl-1-19-multi-teacher-distillation-packing-explained

What is TRL 1.19?

TRL (Transformer Reinforcement Learning) provides trainers for supervised fine-tuning, preference optimisation, reinforcement learning and distillation. Distillation trains a student model to imitate or learn from the outputs of a stronger teacher.

How the new distillation features work

- DistillationTrainer accepts a dictionary of teacher models.
- Each training example has a teacher_id column selecting which teacher scores its completion, the fusion stage of the method called MOPD.
- It also accepts an environment_factory with the same contract as the reinforcement learning trainer, so agents can be distilled against stateful environments.
- Support and an example for Inkling 2 Mini require transformers 5.24.0 or later.

What is new for everyday fine-tuning

- Packed examples stay as separate rows with position IDs restarting per sequence, instead of one flattened row that made scaled dot-product attention quadratic.
- SFTTrainer adds the end-of-turn token to the stop tokens when a checkpoint omits it, so tuned models stop at the end of their turn.
- Assistant-only loss and tool calling work with more chat templates.
- LoRA adapters are merged out of place for vLLM weight sync, which makes QLoRA work in server mode.
- Reward grading runs in a separate process so a pathological answer cannot hang training.

Results reported by the maintainers

- Packed training on H100 with a 1.7B model and SDPA: 2.06 to 1.02 seconds per step. With Flash Attention 2 it is unchanged at about 0.83.
- AsyncGRPO planning and packing on four B300 GPUs: 18 seconds to 1 second per pass, with 28% more trained tokens per second.
- LoRA fine-tuning of Inkling 2 Mini on one H100 for 53 minutes raised held-out clock reading accuracy from 0% to 87.0% in the example task.
- In the QLoRA server-mode fix, a log-probability gap on a small Qwen model fell from 0.19–0.24 to 0.02–0.04.

Limitations and open questions

- Figures come from specific models, GPUs and settings. The clock example is a narrow task, not general quality.
- Teams already using Flash Attention 2 should not expect the packing speedup.
- Upgrading requires peft 0.18.0 and accelerate 1.12.0 or later, and parse_response now needs a prefix argument. Sources were read on 26 November 2026. Products, rules and standards change, so check the linked originals before you rely on any detail.

How to adopt it

1. Pin a branch and upgrade TRL, PEFT, Accelerate and Transformers together.
2. Re-run a short training job and compare loss curves and evaluation scores with your previous version.
3. Check the stop token in your saved generation configuration and test that answers end correctly.
4. If you serve with vLLM 0.34 during training, set --gpu-memory-utilization 0.9 as the documented commands now do.
5. Try multi-teacher distillation only after you have a per-domain evaluation set for each teacher.

What it means for ML engineers

The stop-token and chat-template fixes deserve attention before the headline features: silent template errors can waste a whole training run. Multi-teacher distillation is useful when different specialists are strong in different domains, but each added teacher adds cost, licence conditions and another source of error to audit.

Frequently asked questions

- **What is new in TRL 1.19?** Multi-teacher and agent distillation, faster packed training with SDPA, stop-token handling and many fixes.
- **Does TRL 1.19 support vLLM 0.34?** Yes, the supported range now includes it.
- **Why did my fine-tuned model not stop?** Some chat templates omit the end-of-turn token from the stop list; the new trainer adds it.
- **Is multi-teacher distillation better than one teacher?** The release notes do not claim a general result. Test it on your data.

Related reading

- Harness adapters for agents
- Predicting fine-tuning time
- RAG vs fine-tuning
- Local LLM Practitioner certification

> Related credential: Local LLM Practitioner. Explore the exam facts and syllabus.

Sources and further reading

1. Hugging Face. TRL release v1.19.0. 25 November 2026. github.com/huggingface/trl/releases/tag/v1.19.0

Published by IICTP, an independent technical credential body, on 26 November 2026. The current version of this post is at https://blogs.iictp.org/trl-1-19-multi-teacher-distillation-packing-explained. More guides: blogs.iictp.org.