ICLR 2026

Prompt-MII: Meta-Learning Instruction Induction for LLMs

Emily Xiao · Yixiao Zeng · Ada Chen · Chin-Jou Li · Amanda Bertsch · Graham Neubig

Carnegie Mellon University · Pittsburgh, Pennsylvania · 1900
Language Technologies Institute

1 Motivation

Prompts are the interface for adapting LLMs — yet still **hand-written by humans**: slow, fragile, redone per task.

Today's alternatives — each trades off:

In-context learning
strong F1 — but repeats every example at every query; cost scales with n.

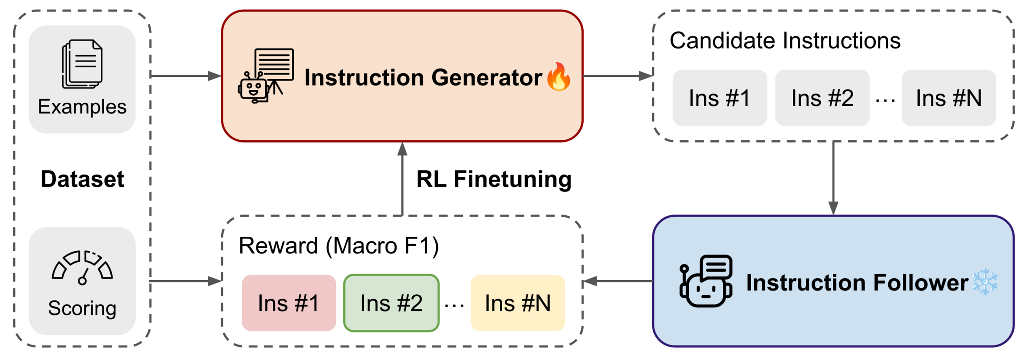
Fine-tuning
cheap at test — but heavy compute, needs weights, often loses to ICL.

Prompt search (APE, GEPA)
automated — but **10^2–10^3 LLM calls *per new task***.

Can one LLM learn to write its own task prompts — for any dataset, in a single forward pass?

2 Method

- ▸ Meta-learn an **Instruction Generator** π_θ via RL across many tasks.
- ▸ Frozen **Follower** LM_{eval} grounds the reward (same base model, not updated).
- ▸ Test time: sample examples → **one forward pass** → reusable instruction.



1 Sample
n examples from task S_i.

2 Generate
π_θ writes instruction I in one pass.

3 Follow
Frozen LM labels m=20 held-out ex.

4 Update
GRPO lifts π_θ toward higher-reward instructions.

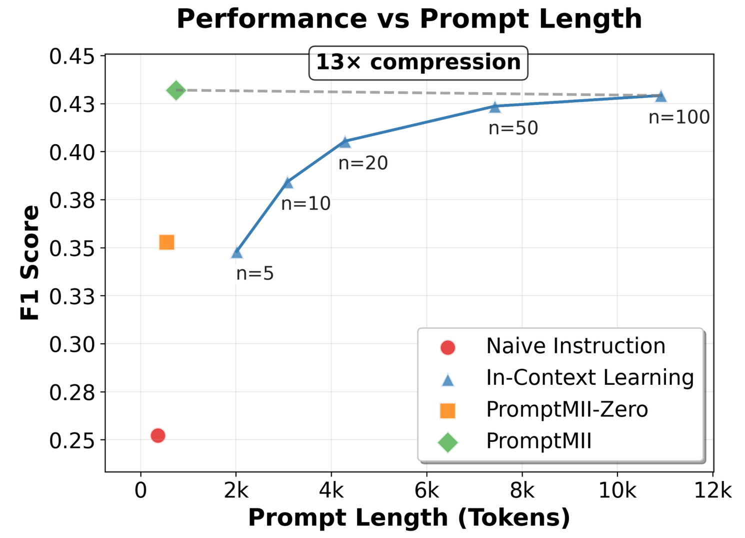
Reward	$R(I) = \text{Macro-F1}(\ \mathrm{LM}_{eval}(I, x_j),\ y_j\ )_{j=1..m}$
Optimizer	**GRPO** · asymmetric clipping · KL removed (DAPO-style — encourages exploration).

3 Key Result

13× fewer tokens than 100-shot ICL at matched F1

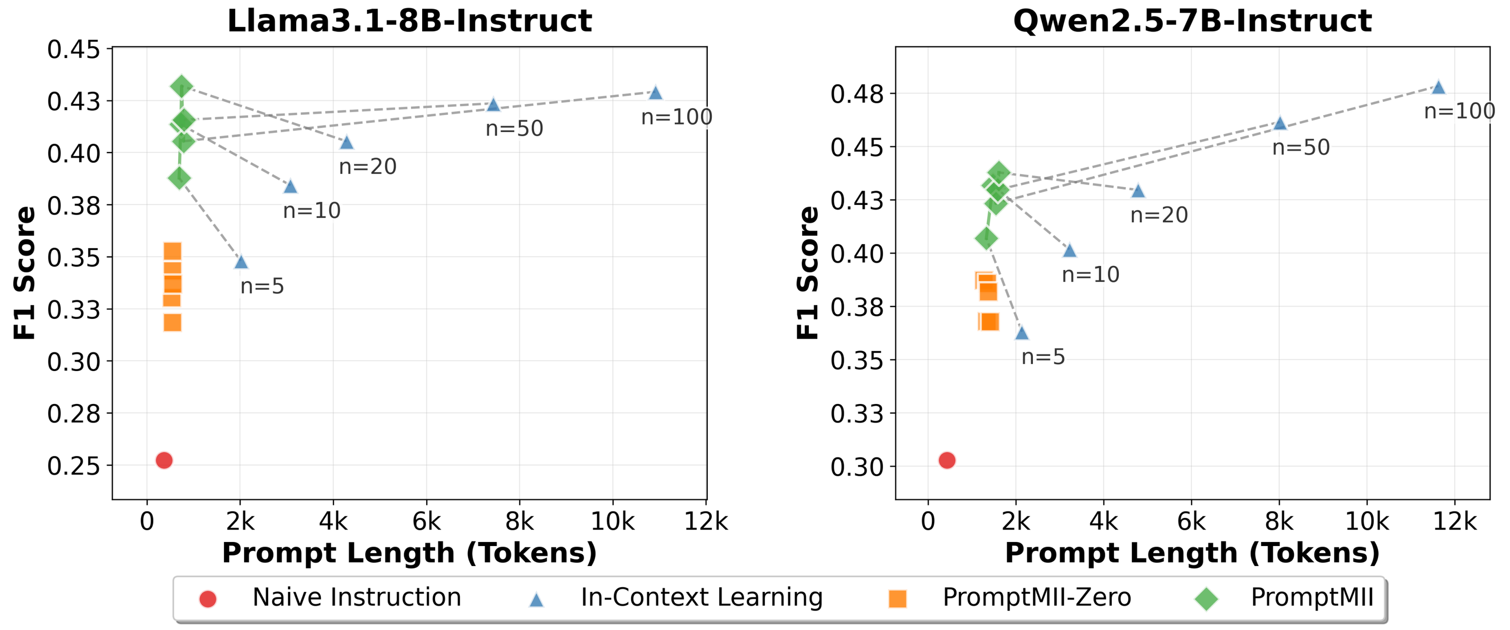
+9 F1 pts over untrained baseline

3,430 training tasks · 90 held-out



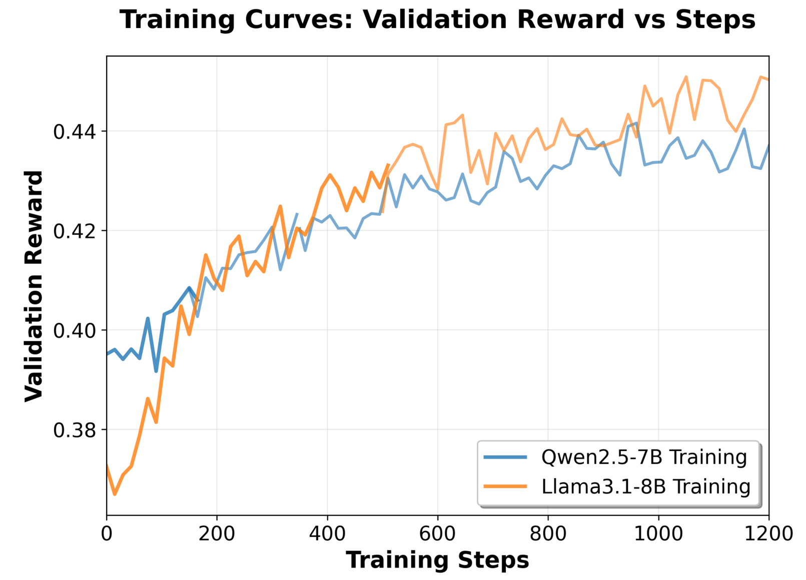
- **Llama-3.1-8B follower** · macro-F1 over 90 held-out tasks.
- Prompt-MII at **n=20** ties 100-shot ICL · **~900** vs. **~11,500 tokens**.
- **~50/50 win rate** vs. ICL across tasks (not significant, Wilcoxon).
- Trained at **4k** ctx · generalises to **32k** at eval.

4 Pareto Frontier



- Prompt-MII (◆): **upper-left Pareto frontier**, both backbones.
- **Higher F1 · fewer tokens** than every baseline.
- Holds at **every prompt-length budget**.

5 Training Dynamics



- Validation reward rises steadily on both backbones throughout training.
- Gains measured on **381 held-out classification tasks**, separate from the 90 eval tasks.
- First evidence that **one-pass instruction induction** is an RL-learnable skill.

6 vs. APE and GEPA

Method	Llama-3.1-8B		Qwen2.5-7B	
	n=50	n=100	n=50	n=100
Naive	.253	.253	.303	.303
APE	.278	.288	.358	.356
GEPA	.296	.299	.346	.347
Prompt-MII	**.416**	**.405**	**.432**	**.424**

- **+0.08–0.14 F1** over APE / GEPA, both backbones.
- Trained once · APE / GEPA re-search *per task*.

Test-time LLM calls per task

7 Cross-Model Transfer

8B Llama Prompt-MII > **405B** Llama Zero-shot

- Trained **8B** writer > zero-shot **405B** writer.
- Training the writer beats scaling it.

8 Generator Outcome

Same **n=10 examples** · Llama-3.1-8B follower · 4-label forum-post task — what each generator writes:

Prompt-MII-Zero (untrained)

```
Classify the input text as one of the following labels: 1; 0; 2; 3.
• The presence of a question or request for help, often indicated by phrases such as "I
  need"…
• The tone and language used (e.g. polite language; uncertainty; or a sense of seeking
  guidance).
• The content of the text, which may relate to a specific topic (e.g. computer hardware;
  medical careers; cryptocurrency).
Respond with only the label name – one of: 1; 0; 2; 3.
```

- **Vague** heuristics.
- Domain **not inferred**.
- **Generic** template.

F1 **0.241**

Prompt-MII (RL-trained)

```
Classify each input into one of the following categories based on content and purpose:
• Label 0: asking for advice, guidance, or recommendations on building / upgrading a
  computer, purchasing components, or troubleshooting hardware issues.
• Label 3: unrelated to computers – other topics such as business, finance, or
  cryptocurrency.
• Label 2: advice on non-computer topics – education, career, or personal development.
• Label 1: does not fit any of the above; unclear or ambiguous inputs.
Respond with the corresponding label (0, 1, 2, or 3) only.
```

- **Domain inferred**: computer-hardware forum.
- **Per-label boundaries.**
- **Generalisable** pattern.

F1 **0.829**

★ Conclusions

- One policy, **one forward pass**, no per-task search.
- **3–13× shorter prompts** · prefix-cacheable.
- Natural-language output · **transfers** to any follower LLM.

Code | Checkpoints | Paper