Multi-Task Transfer Matters During Instruction-Tuning

David Mueller ● Nicholas Andrews ● Mark Dredze

Johns Hopkins University ○ Department of Computer Science ○ Baltimore, MD

Although the goal of instruction-tuning is to improve in-context learning (ICL) ability, it induces a multi-task optimization problem under which several hundred tasks are jointly optimized.

Multi-task training is known to lead to negative transfer when improperly optimized, and can cause catastrophic forgetting in large, pre-trained models, which may harm ICL performance. We ask...

- **Does the impact of multi-task learning on model *transfer* matter for in-context learning?**
- **Can we improve ICL by mitigating the negative effects of MTL?**

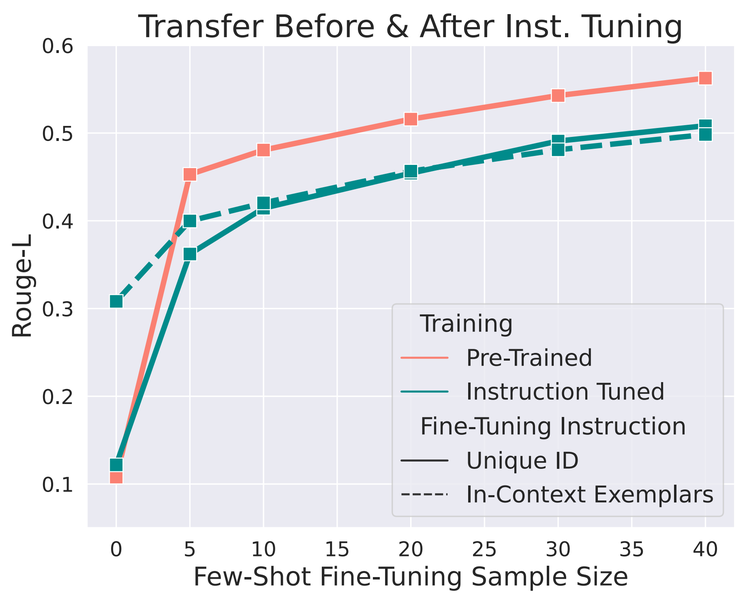
Spoiler: We find that the answer is yes to both questions!

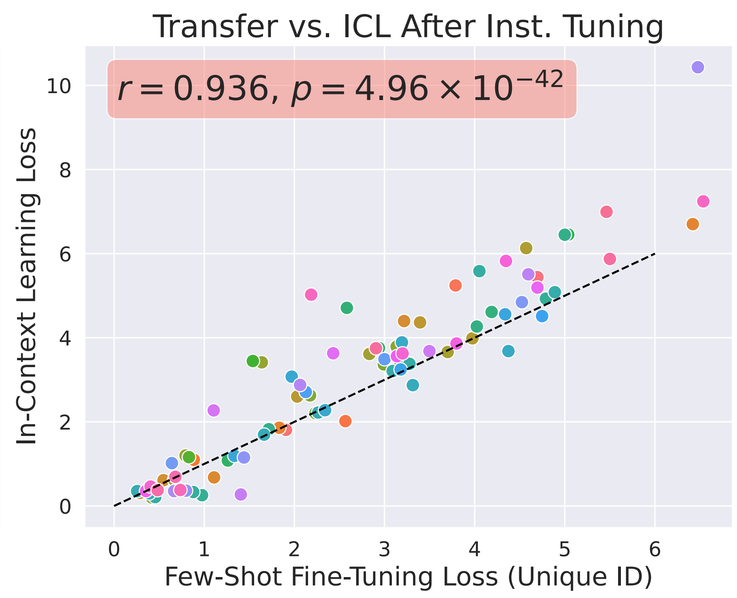
Instruction-tuning can cause *catastrophic forgetting* from the perspective of *parameter transfer* (i.e. transfer via fine-tuning).

...and, parameter transfer is highly correlated with in-context learning performance on unseen tasks.

How well a model fine-tunes to a task may be indicative of how successfully we can learn that task in-context... and MTL impacts fine-tuning transfer!

Instruction-Tuning causes catastrophic forgetting, which may impact ICL to unseen tasks.

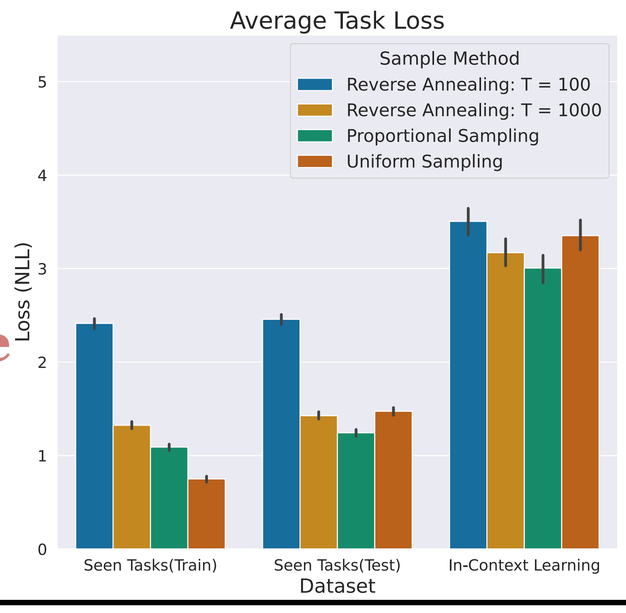




In-Context Learning is impacted by different MTL training schemes!

Does it matter how we fit the MTL objective of instruction-tuning, or just how many instructions we see? We instruction-tune with different MTL training methods...

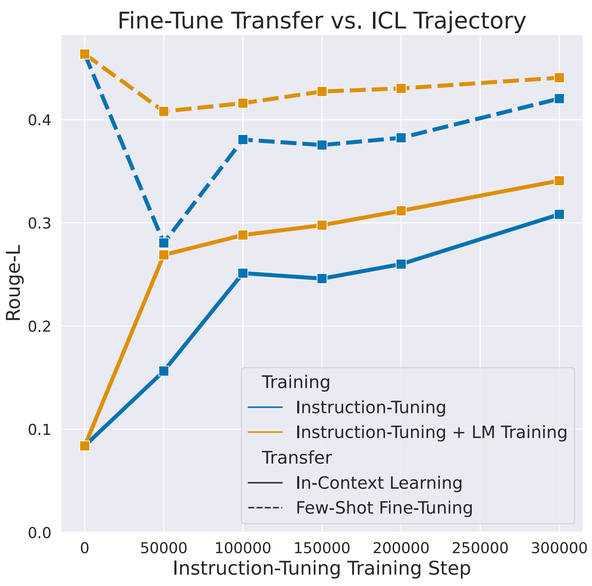
The method which yields the best MTL generalization also yields the best ICL performance, despite all methods seeing the same number of instructions!



Mitigating catastrophic forgetting improves ICL performance!

If we attempt to reduce catastrophic forgetting, we can improve transfer to unseen tasks via fine-tuning. Does this impact ICL?

Continual pre-training improves ICL performance, despite only adding previously seen data. Fine-tuning transfer matters for ICL!



Conclusion

There is a connection between a model's ability to transfer to an unseen task (via fine-tuning) and learn that task in-context after instruction-tuning. By improving parameter transfer during instruction-tuning, we can improve ICL.