Few-shot **Fine-tuning** vs. **In-context Learning**

A Fair Comparison and Evaluation


Marius Mosbach[1]

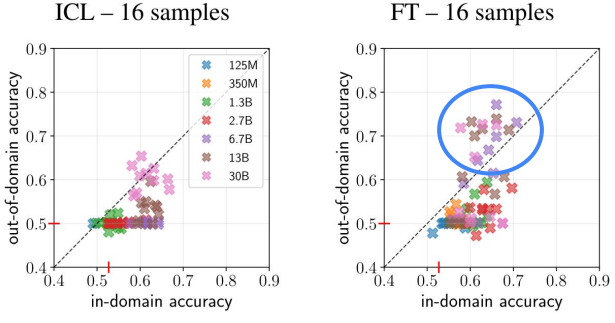

Tiago Pimentel[2]


Shauli Ravfogel[3]

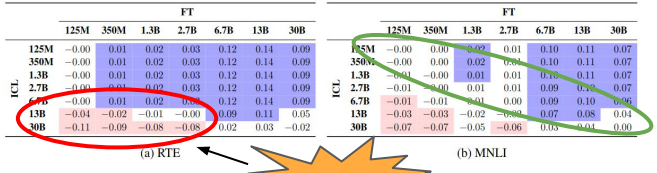

Dietrich Klakow[1]


Yanai Elazar[4,5]

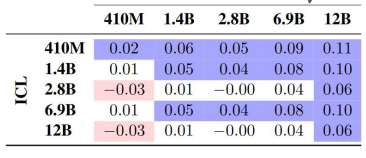
1 Test generalization for models adapted via **fine-tuning** (FT) and **in-context learning** (ICL)



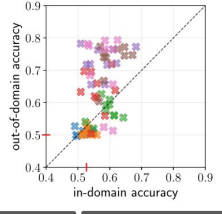
2 Findings **generalize** to other models (Pythia) and FT via LoRA



Unfair!



PBFT + LoRA



Significance tests: ICL > FT, FT > ICL

3 **High-level comparison:**

	FT	ICL
Users	Experts	Experts & Non-experts
Interaction	Pre-defined	Textual
Reusability	Medium	High
Applicability to low-resource languages	High	Limited
Requires training	Yes	No
Number of samples	Unlimited	0 to ?
Variance	High	High
Well understood	No	No

Fine-tuned models can generalize well out-of-domain

Approaches should be compared using models of the same size

Robust task adaptation remains a challenge!

uds-lsv/llmft

[1] Saarland University, [2] University of Cambridge, [3] Bar Ilan University, [4] Allen Institute for Artificial Intelligence, [5] University of Washington

61 ACL 2023