

# Improving In-Context Few-Shot Learning via Self-Supervised Training

Mingda Chen<sup>1</sup>, Jingfei Du<sup>2</sup>, Ramakanth Pasunuru<sup>2</sup>, Todor Mihaylov<sup>2</sup>, Srini Iyer<sup>2</sup>, Veselin Stoyanov<sup>2</sup>, Zornitsa Kozareva<sup>2</sup>

<sup>1</sup>Toyota Technological Institute at Chicago <sup>2</sup>Meta AI

## In-Context Few-Shot Learning

Solve **unseen tasks** at inference time while **forgoing any weight updates**

Task demo. **Input:** Context word: fit. Question: The trophy doesn't fit into the brown suitcase because \_\_ is too large.  
**Output:** trophy.  
... (extra input-output pairs for the same task)

**Input:** Context word: water. Question: I poured water from the bottle into the cup until the \_\_ was empty.

**Output:** **bottle**

Model prediction

## Prior Work

Language Model Pre-Training

Extra Intermediate Fine-Tuning Step

Fine-Tuning on Human-Annotated Datasets

Evaluation: In-Context Few-Shot Learning

## Our Work

Can **self-supervised tasks** be used in the intermediate fine-tuning steps?

## Self-Supervised Tasks

### Original Raw Text:

Natural language processing is a subfield of computer science. The goal is a computer capable of "understanding" the contents of documents.

### Next Sentence Generation (NSG)

**Input:** Natural language processing is a subfield of computer science.

**Output:** The goal is a computer capable of "understanding" the contents of documents.

### Masked Word Prediction (MWP)

**Input:** Natural language processing is a subfield of \_\_. The goal is a computer capable of "understanding" the contents of documents.

**Output:** Computer Science

### Last Phrase Prediction (LPP)

**Input:** Natural language processing is a subfield of computer science. Question: The goal is a computer capable of "understanding"?

**Output:** the contents of documents.

**Input:** Natural language processing is a subfield of computer science. Question: The goal is a computer capable of "understanding"? Answer: the development of new models.

**Output:** False.

### Classification (CL)

For example, when input consists of sentences from multiple documents:

**Input:** Natural language processing is a subfield of computer science. Computer vision deals with ...

**Output:** False

## Experiments

### Setup

- ❖ We constructed data from the RoBERTa training corpus.
- ❖ The final training data contains approximately 1 million instances with 250k training instances per task.
- ❖ We evaluate models on five tasks from SuperGLUE and eight tasks from Natural-Instructions.
- ❖ Baseline: ExtraLM: Perform additional LM pre-training on the portion of the original raw text used in our self-supervised training.
- ❖ Baseline: CrossTask: Using human-annotated datasets in the intermedia fine-tuning step.

### Results

	BoolQ	MultiRC	COPA	RTE	CB	Avg.
LM	48.6	5.5/53.7	83.4	51.9	53.6	51.8
ExtraLM	49.6	4.9/54.8	82.6	52.9	51.4	51.7
CrossTask	53.4	1.2/57.2	76.2	<b>54.3</b>	44.6	49.6
SelfSup	<b>61.7</b>	<b>5.2/62.1</b>	<b>84.0</b>	53.1	<b>54.3</b>	<b>55.6</b>

### SuperGLUE results

	QG	AG	MM	VF	Avg.
GPT3	43.0	50.0	70.0	32.0	48.8
LM	40.9	32.5	74.0	27.8	43.8
ExtraLM	41.1	32.7	<b>75.9</b>	25.2	43.7
CrossTask	38.1	<b>41.6</b>	69.2	23.0	42.9
SelfSup	<b>43.9</b>	37.5	72.3	<b>28.6</b>	<b>45.5</b>

### Natural-Instructions results