Super-NaturalInstructions: Generalization via Declarative Instructions on 1600+ NLP Tasks
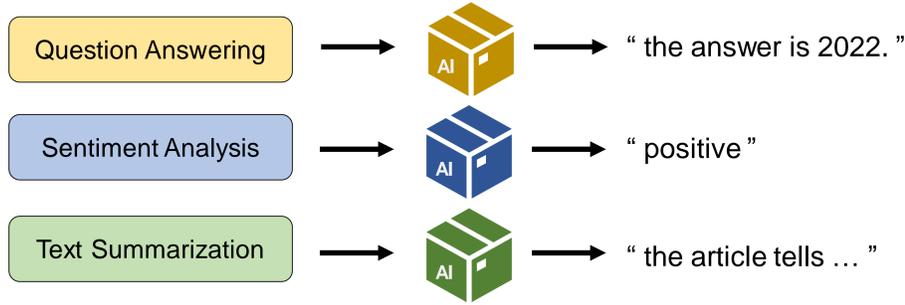
🐦 @yizhongwyz

✉ yizhongw@cs.washington.edu

🌐 Website

Yizhong Wang*, Swaroop Mishra*, Pegah Alipoormolabashi, Yeganeh Kordi, Chitta Baral, Yejin Choi, Noah A. Smith, Hannaneh Hajishirzi, and Daniel Khashabi, and others
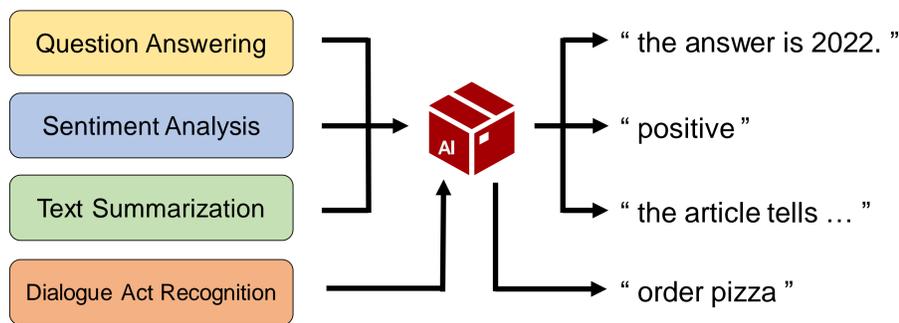
UWNLP   Ai2

Background

❖ Task-specific Models

Question Answering → AI → " the answer is 2022. "

Sentiment Analysis → AI → " positive "

Text Summarization → AI → " the article tells … "

❖ General-purpose Model

Question Answering
Sentiment Analysis
Text Summarization
Dialogue Act Recognition
→ AI →
" the answer is 2022. "
" positive "
" the article tells … "
" order pizza "

Large pretrained language models can solve various types of tasks by following in-context instructions.

Instructing Example

Task Instruction

Definition

"... Given an utterance and recent dialogue context containing past 3 utterances (wherever available), output 'Yes' if the utterance contains the small-talk strategy, otherwise output 'No'. Small-talk is a cooperative negotiation strategy. It is used for discussing topics apart from the negotiation, to build a rapport with the opponent."

Positive Examples

- **Input:** "Context: … *'That's fantastic, I'm glad we came to something we both agree with.'* Utterance: *'Me too. I hope you have a wonderful camping trip.'*"
- **Output:** "Yes"
- **Explanation:** "The participant engages in small talk when wishing their opponent to have a wonderful trip."

Negative Examples

- **Input:** "Context: … *'Sounds good, I need food the most, what is your most needed item?!'* Utterance: *'My item is food too'.*"
- **Output:** "Yes"
- **Explanation:** "The utterance only takes the negotiation forward and there is no side talk. Hence, the correct answer is 'No'."

Tk-Instruct
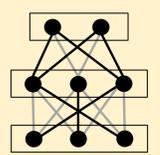
Evaluation Instances

- **Input:** "Context: … *'I am excited to spend time with everyone from camp!'* Utterance: *'That's awesome! I really love being out here with my son. Do you think you could spare some food?'*"
- **Expected Output:** "Yes"

Super-Natural Instructions

*We collected a diverse set of **1,616** NLP tasks and annotated their natural language **instructions**. (see example 👉)*

❖ **How are they collected?**
- ➤ Data was contributed by 88 NLP practitioners from the community.
- ➤ Instructions were carefully written by these contributors and then reviewed by experts.
- ➤ Multiple iterations of editing and review were done via GitHub to ensure quality.

❖ **Why is this dataset unique?**
- ➤ Instructions are declarative and informative!
- ➤ Diversity: **76** broad categories (see comparison 👇)

(a) SUP-NATINST (this work)

(b) NATINST

(c) PROMPTSOURCE (T0 subset)
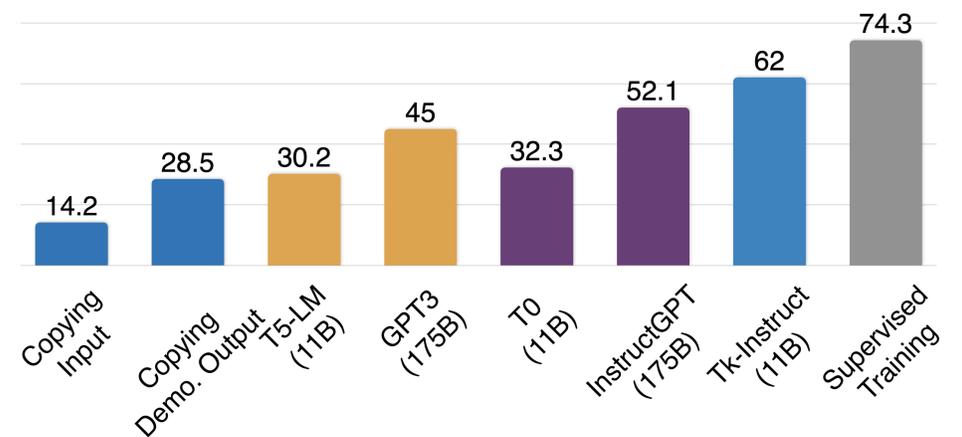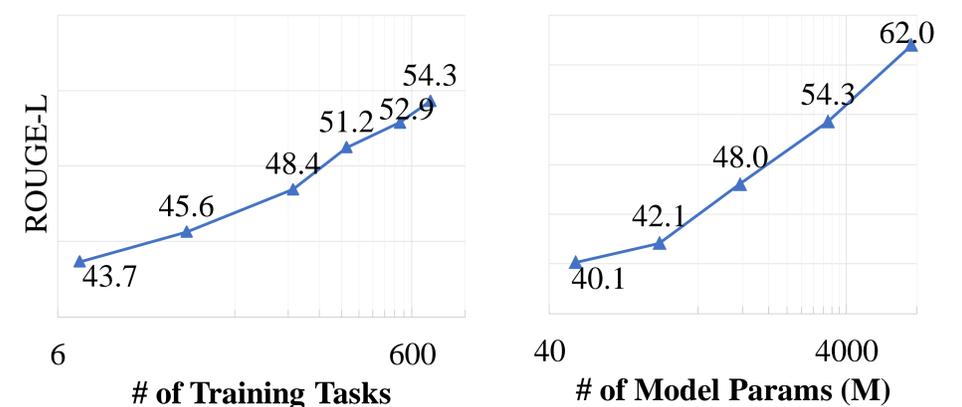
(d) FLAN

(e) INSTRUCTGPT
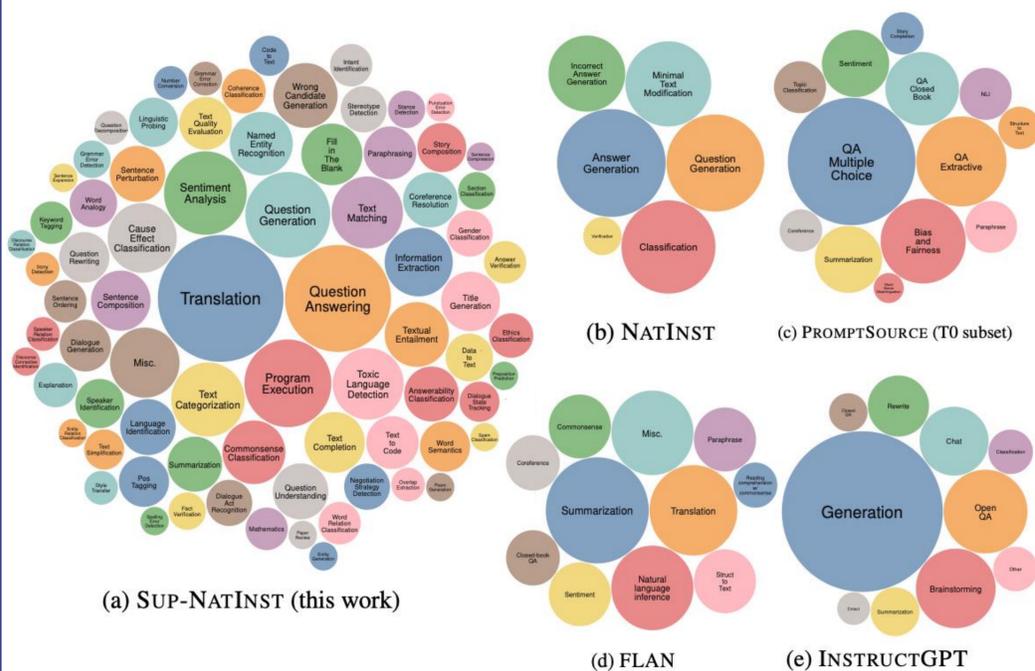
Tk-Instruct

Instruction tuning of T5 model (11B) on our data enables better generalization to unseen tasks than GPT3 (175B).

❖ **Eval setup for cross-task generalization**
- ➤ 12 manually picked evaluation categories.
- ➤ English track: 119 eval tasks, 757 training tasks.
- ➤ Cross-lingual track: 35 eval tasks, 1271 training tasks.

Copying Input	Copying Demo. Output	T5-LM (11B)	GPT3 (175B)	T0 (11B)	InstructGPT (175B)	Tk-Instruct (11B)	Supervised Training
14.2	28.5	30.2	45	32.3	52.1	62	74.3

❖ **Scaling trend of generalization**

ROUGE-L: 43.7, 45.6, 48.4, 51.2, 52.9, 54.3
# of Training Tasks: 6 — 600

40.1, 42.1, 48.0, 54.3, 62.0
# of Model Params (M): 40 — 4000

❖ **Check our paper for more ablations!**
- ➤ Definition and in-context examples are complementary.
- ➤ A large number of training instances are not necessary.

References:
[1] Mishra et al. "Cross-Task Generalization via Natural Language Crowdsourcing Instructions". ACL 2022.
[2] Sanh et al. "Multitask Prompted Training Enables Zero-Shot Task Generalization". ICLR 2022.
[3] Bach et al. "PromptSource: An Integrated Development Environment and Repository for Natural Language Prompts". ACL 2022.
[4] Wei et. al. "Finetuned Language Models are Zero-Shot Learners." ICLR 2022.
[5] Ouyang et al. "Training Language Models to Follow Instructions with Human Feedback". arXiv preprint 2022.