Instruction Tuning

- We want to optimize models for P(answer | prompt), but they're learned on a basic language modeling objective P(word | context)
- One solution: fine-tune these models to do what we care about (question answering, classification, …)
- Two main ways of doing this in 2023:
  - **Instruction tuning:** supervised fine-tuning on data derived from many NLP tasks
  - **Reinforcement learning from human feedback (RLHF):** RL to improve human judgments of how good the outputs are

Task Generalization: T0

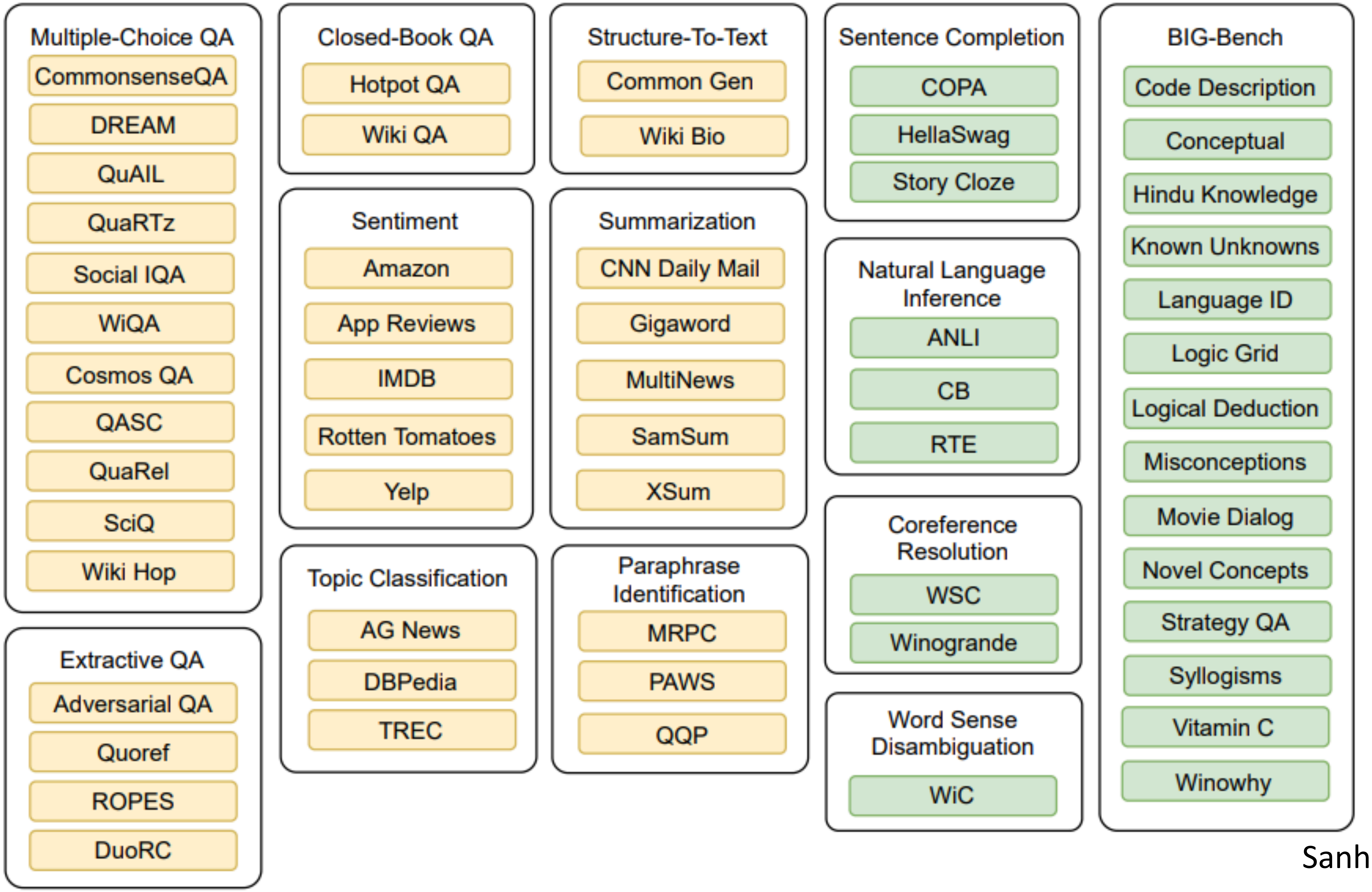
- T0: tries to deliver on the goal of T5 and do many tasks with one model
- **Crowdsourced prompts**: instructions for how to do the tasks

Summarization
The picture appeared on the wall of a Poundland store on Whymark Avenue [...] How would you rephrase that in a few words?
Paraphrase identification
"How is air traffic controlled?" "How do you become an air traffic controller?" Pick one: these questions are duplicates or not duplicates.
Question answering
I know that the answer to "What team did the Panthers defeat?" is in "The Panthers finished the regular season [...]". Can you tell me what it is?
Multi-task training
Zero-shot generalization
Natural language inference
Suppose "The banker contacted the professors and the athlete". Can we infer that "The banker contacted the professors"?
T0
Graffiti artist Banksy is believed to be behind [...]
Not duplicates
Arizona Cardinals
Yes

Sanh et al. (2021)

Task Generalization: T0

- Pre-train: T5 task
- Train: a collection of tasks with prompts. **This uses existing training data**
- Test: a new task specified only by a new prompt. **No training data in this task**



Multiple-Choice QA
CommonsenseQA
DREAM
QuAIL
QuaRTz
Social IQA
WiQA
Cosmos QA
QASC
QuaRel
SciQ
Wiki Hop

Extractive QA
Adversarial QA
Quoref
ROPES
DuoRC

Closed-Book QA
Hotpot QA
Wiki QA

Sentiment
Amazon
App Reviews
IMDB
Rotten Tomatoes
Yelp

Topic Classification
AG News
DBPedia
TREC

Structure-To-Text
Common Gen
Wiki Bio

Summarization
CNN Daily Mail
Gigaword
MultiNews
SamSum
XSum

Paraphrase Identification
MRPC
PAWS
QQP

Sentence Completion
COPA
HellaSwag
Story Cloze

Natural Language Inference
ANLI
CB
RTE

Coreference Resolution
WSC
Winogrande

Word Sense Disambiguation
WiC

BIG-Bench
Code Description
Conceptual
Hindu Knowledge
Known Unknowns
Language ID
Logic Grid
Logical Deduction
Misconceptions
Movie Dialog
Novel Concepts
Strategy QA
Syllogisms
Vitamin C
Winowhy

Sanh et al. (2021)

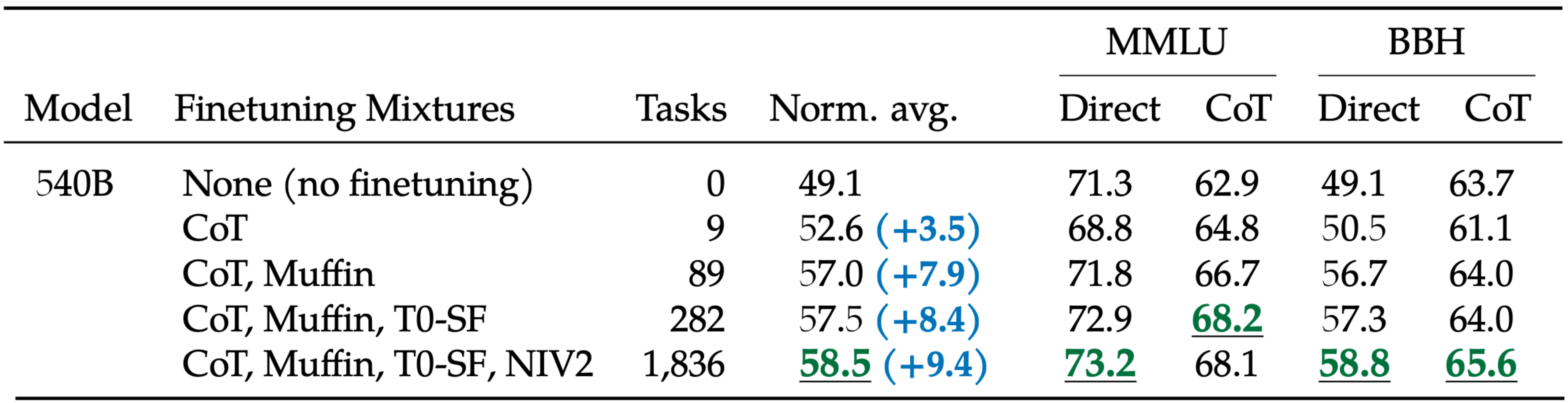
Flan-PaLM

- Flan-PaLM (October 20, 2022): 1800 tasks, 540B parameter model fine-tuned on many tasks after pre-training

Instruction finetuning
Please answer the following question.
What is the boiling point of Nitrogen?
-320.4F
Chain-of-thought finetuning
Answer the following question by reasoning step-by-step.
The cafeteria had 23 apples. If they used 20 for lunch and bought 6 more, how many apples do they have?
Language model
The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9.
Multi-task instruction finetuning (1.8K tasks)
Inference: generalization to unseen tasks
Q: Can Geoffrey Hinton have a conversation with George Washington?
Give the rationale before answering.
Geoffrey Hinton is a British-Canadian computer scientist born in 1947. George Washington died in 1799. Thus, they could not have had a conversation together. So the answer is "no".

Chung et al. (2022)

Flan-PaLM: Results



				MMLU		BBH	
Model	Finetuning Mixtures	Tasks	Norm. avg.	Direct	CoT	Direct	CoT
540B	None (no finetuning)	0	49.1	71.3	62.9	49.1	63.7
	CoT	9	52.6 (+3.5)	68.8	64.8	50.5	61.1
	CoT, Muffin	89	57.0 (+7.9)	71.8	66.7	56.7	64.0
	CoT, Muffin, T0-SF	282	57.5 (+8.4)	72.9	**68.2**	57.3	64.0
	CoT, Muffin, T0-SF, NIV2	1,836	**58.5** (+9.4)	**73.2**	68.1	**58.8**	**65.6**

- Human performance estimates are ~80 on Big-Bench (BBH)
- MMLU: multiple-choice test questions drawn from many disciplines
- Note: smaller 11B versions of these models are released (Flan-T5-11B); still a good choice for many tasks!

Chung et al. (2022)