

API Pack: A Massive Multi-Programming Language Dataset for API Call Generation

Zhen Guo
MIT EECS

Adriana Meza Soria*
MIT-IBM Watson AI lab

Wei Sun
IBM Research

Yikang Shen
MIT-IBM Watson AI lab

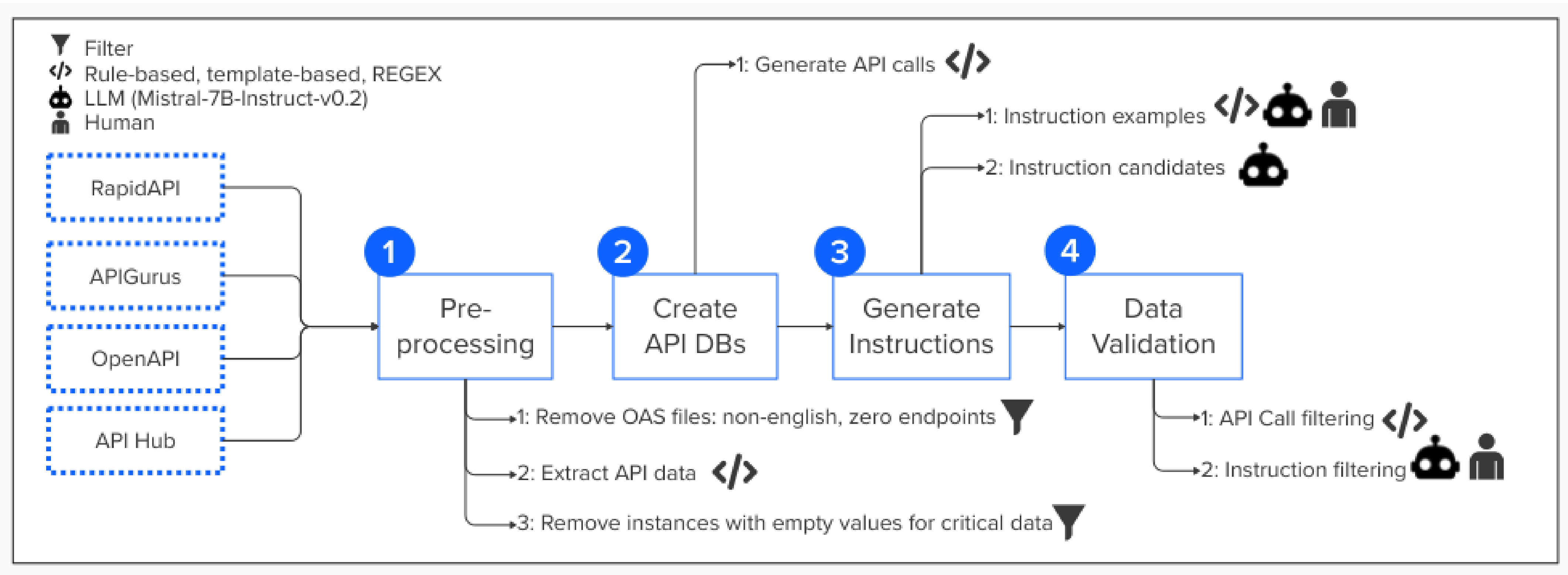
Rameswar Panda
MIT-IBM Watson AI lab

Why are API and tool datasets important?

- API datasets are important to improve the ability of large language models (LLMs) to generate API calls.
- Accurate API call generation is a key capability required in LLMs to create LLM powered systems able to interact with API ecosystems and complete end-to-end complex operations.
- Previous work focused on API intent detection (finding the right API for a task), but generating API call code in multiple programming languages has received less attention.
- Moreover, most publicly available API and tool datasets are relatively small.

Feature	API Pack (this work)	APIBench (Gorilla)	ToolBench	ToolBench (ToolLLM)	API Bank	ToolAlpaca	ToolFormer
API call intent detection?	✓	✓	✓	✓	✓	✓	✓
API call code generation?	✓	✓	✓	✗	✗	✗	✗
Multi-programming language?	✓(10)	✗(Python)	✓(Curl, Python)	✗	✗	✗	✗
Multi-API call scenario?	✗	✗	✓	✓	✓	✓	✓
Data generation method	custom	self-instruct	self-instruct	custom	custom	custom	custom
# of Sources	4	3	8	1	53	/	5
# of APIs / Tools	11,213	1,645	8	16,464	53	400	5
# of API calls	1,128,599	16,450	/	37,204	568	3,938	9,400
# of Instances	1,128,599	16,450	2,746	12,657	264	3,938	22,453

API Pack curation



Experimental design

Evaluation framework

- Scenario 1: Generalization to new instructions (seen APIs and endpoints)
- Scenario 2: Generalization to new endpoints (seen APIs and new endpoints)
- Scenario 3: Generalization to new APIs (unseen APIs and endpoints)

Experimental settings

- Fine-tuning
- Cross-language generalization
- Scaling

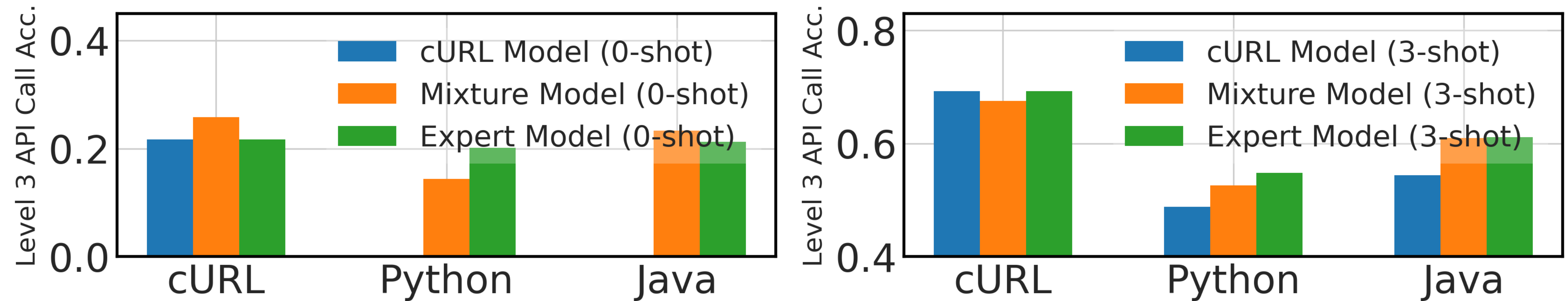
Finetuning results

Fine-tuning CodeLlama-13B on 20,000 Python API Pack instances outperforms GPT-3.5 by 10% and GPT-4 by 5% on new API calls not seen during training.

			Evaluation Accuracy (%)					
			Scenario 1 Seen APIs and endpoints		Scenario 2 Seen APIs and new endpoints		Scenario 3 New APIs and endpoints	
Model	Fine-tuning	Testing	Intent	API Call	Intent	API Call	Intent	API Call
Mistral-7b	0-shot	0-shot	17.2	10.9	14.1	11.4	14.3	11.2
		3-shot (retrieval)	42.0	29.7	35.4	28.7	39.1	29.1
	3-shot	0-shot	40.5	28.5	24.0	18.3	15.2	12.1
		3-shot (retrieval)	64.1	55.4	49.1	42.8	50.8	42.5
CodeLlama-7b	0-shot	0-shot	8.1	6.1	10.0	7.0	11.0	7.8
		3-shot (retrieval)	52.6	42.6	43.6	35.9	50.2	40.1
	3-shot	0-shot	12.1	9.3	13.7	10.2	16.8	13.0
		3-shot (retrieval)	60.6	52.7	54.1	47.3	55.9	49.1
Llama-2-13b	0-shot	0-shot	9.4	6.2	11.6	9.0	10.9	8.4
		3-shot (retrieval)	44.5	33.9	45.4	35.6	46.7	39.1
	3-shot	0-shot	15.7	10.2	14.0	11.2	11.7	9.6
		3-shot (retrieval)	59.5	51.5	50.8	44.3	52.7	44.2
CodeLlama-13b	0-shot	0-shot	9.8	6.8	10.8	8.1	12.1	8.5
		3-shot (retrieval)	55.6	44.4	50.6	43.3	52.3	44.1
	3-shot	0-shot	14.4	10.3	15.9	13.3	14.2	8.9
		3-shot (retrieval)	63.5	55.5	56.8	51.4	56.1	49.5
	none	0-shot	0.1	0.0	0.2	0.0	0.1	0.0
		3-shot (retrieval)	49.2	46.2	36.3	34.4	40.7	38.5
gpt-3.5-1106	none	0-shot	-	-	-	-	1.0	0.7
		3-shot (retrieval)	-	-	-	-	47.2	39.5
gpt-4-1106	none	0-shot	-	-	-	-	0.2	0.1
		3-shot (retrieval)	-	-	-	-	53.5	44.3

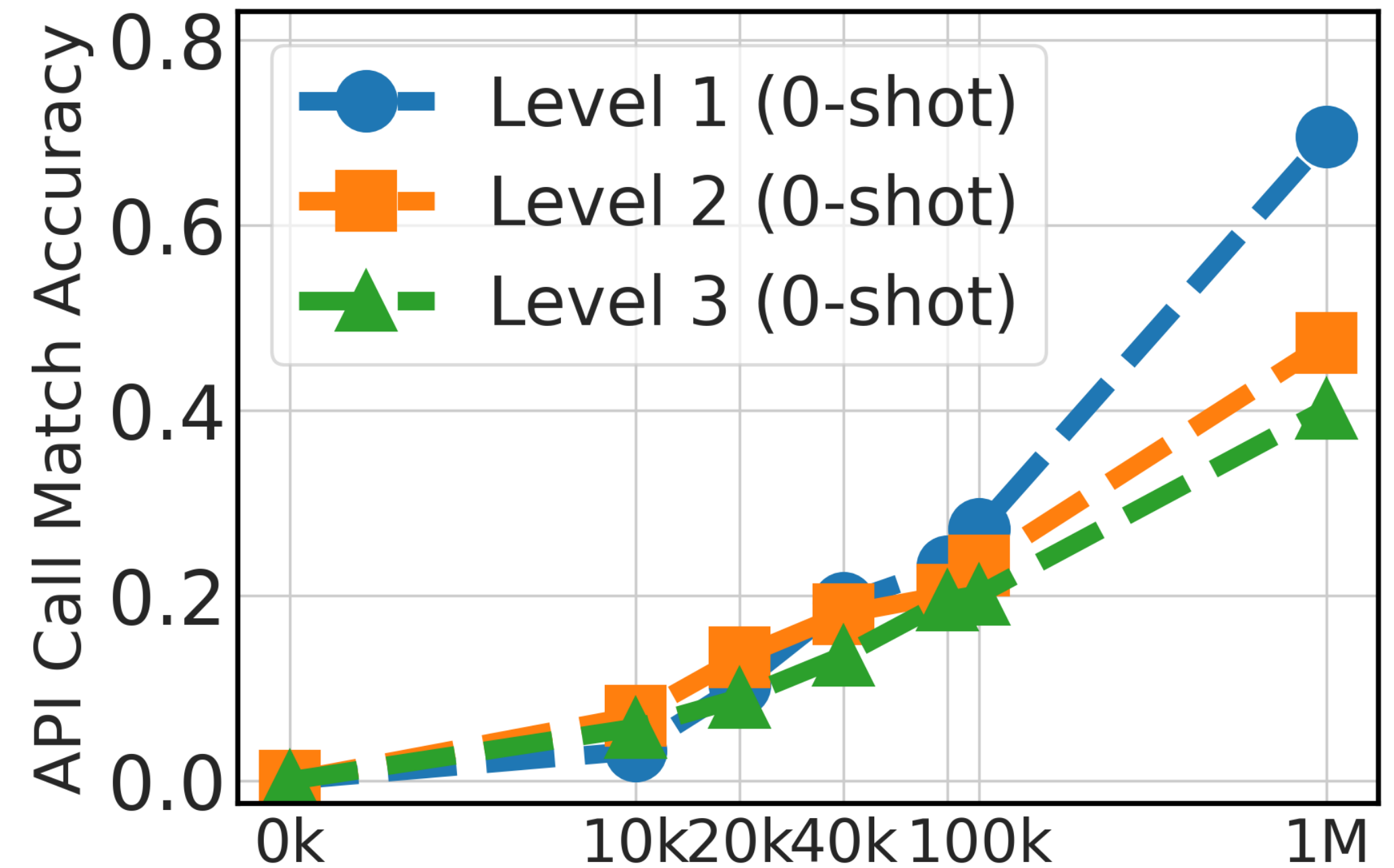
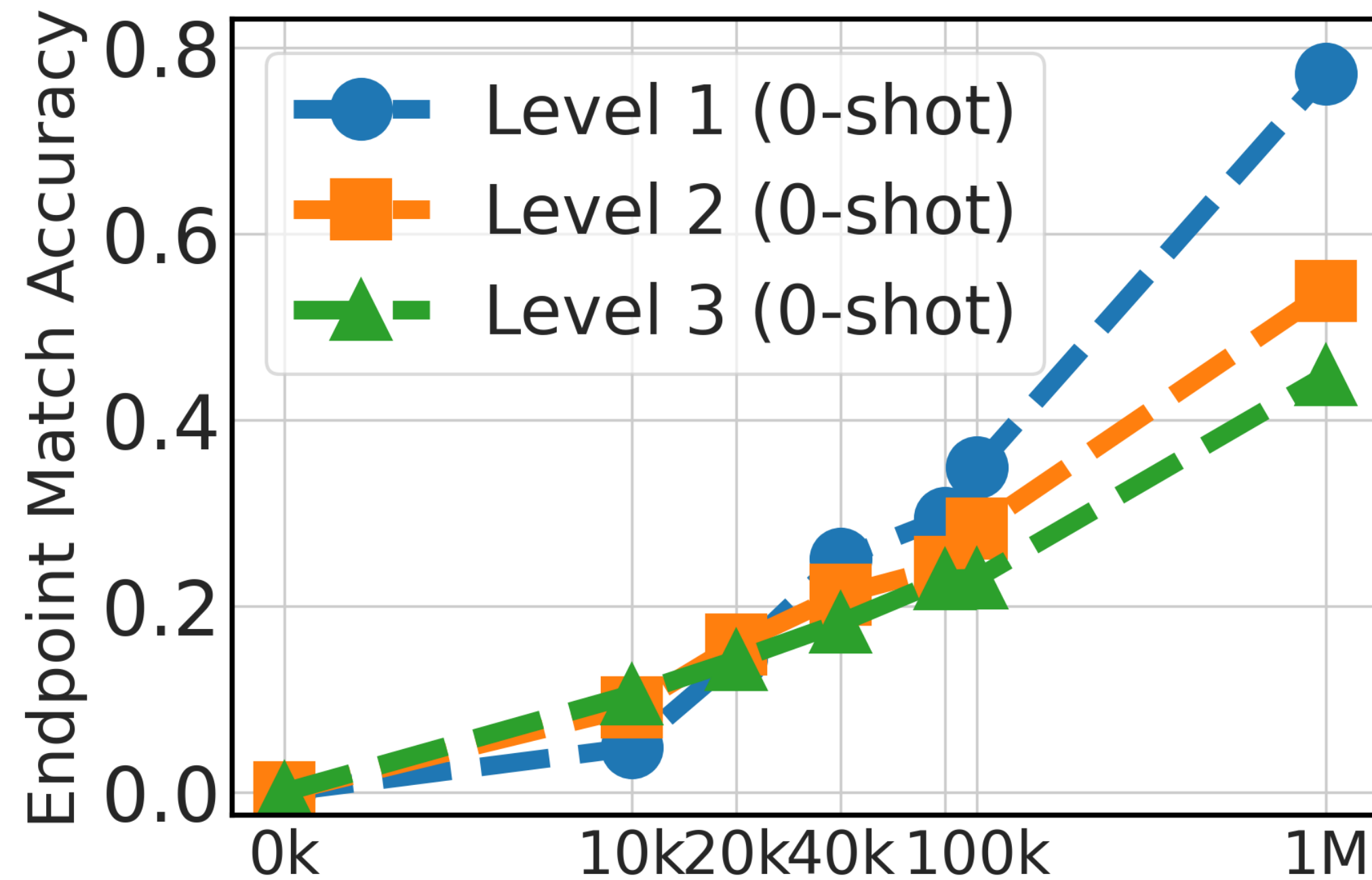


Key findings about cross-language generalization



Cross-language API call generation can be enabled by a large volume of data in one programming language along with smaller amounts of data in others.

Key findings about the scaling experiment



Increasing fine-tuning data from zero to one million instances improves generalization to new APIs, demonstrating the clear benefits of larger datasets for API generalization.

API Pack Limitations and Future Work

Limitations

- API Pack was not designed for scenarios involving multiple API calls, and its instructions require explicit API name specification, which may limit its applicability in complex software development scenarios.
- The API Hubs we sourced data from are all public and operate under a CC0-1.0 license. However, the availability of API-level license information is limited, which may restrict the use of API Pack in proprietary development contexts.

Future Work

- We plan to expand the dataset to support multi-API scenarios.
- We aim to leverage an automatic approach to create synthetic instructions that map natural language to an API call without requiring specification of the API to use.
- We plan to increase programming language diversity.

Thanks



Adriana Meza Soria

Email: adriana.meza.soria@ibm.com

Website: <https://ameza13.github.io/adriana-meza-soria/>

LinkedIn: [adriana-meza-soria-52799961](https://www.linkedin.com/in/adriana-meza-soria-52799961)