Large Language Models

Understanding In-Context Learning

Sharif University of Technology
Soleymani
Fall 2023

In-Context Learning

In-Context Learning (ICL)

- Given a **prompt** including:
  - An optional description of the task
  - Few-shot training examples in a prompt format demonstrating the task
  - The test input

"Apple → Red, Lime → Green, Corn →"

"Albert Einstein was German \n Mahatma Gandhi was Indian \n Marie Curie was"

Brown et al., Language Models are Few-Shot Learners, 2020

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
Translate English to French:     ← task description
cheese =>                        ← prompt
```

One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

```
Translate English to French:     ← task description
sea otter => loutre de mer       ← example
cheese =>                        ← prompt
```

Few-shot

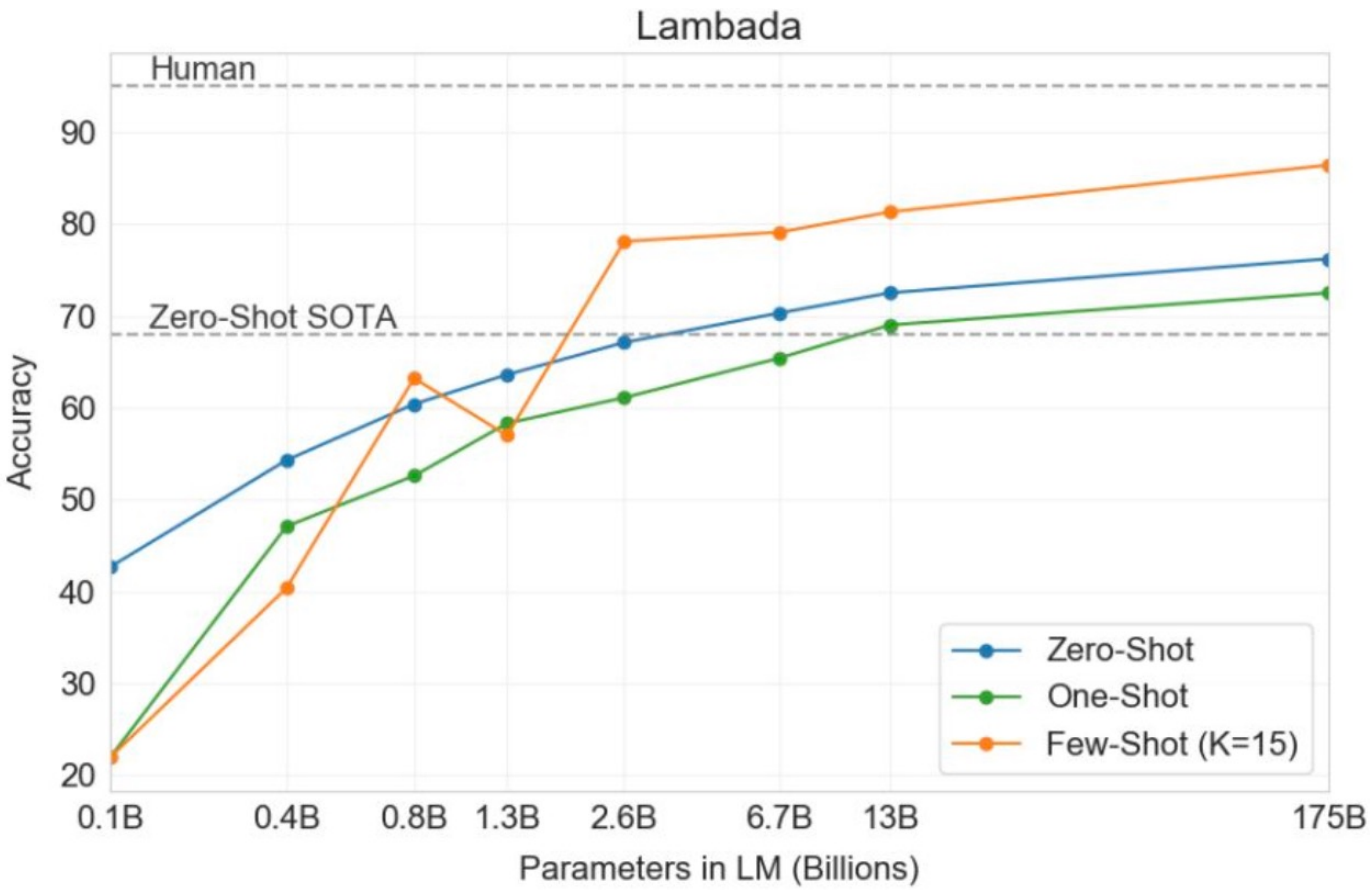
In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
Translate English to French:     ← task description
sea otter => loutre de mer       ← examples
peppermint => menthe poivrée     ←
plush girafe => girafe peluche   ←
cheese =>                        ← prompt
```

What can ICL do?

- No parameter tuning is required
- Only need few examples for downstream tasks
- GPT-3 improved
  SOTA on LAMBADA by 18%!

Works like magic!



Brown et al., Language Models are Few-Shot Learners, 2020

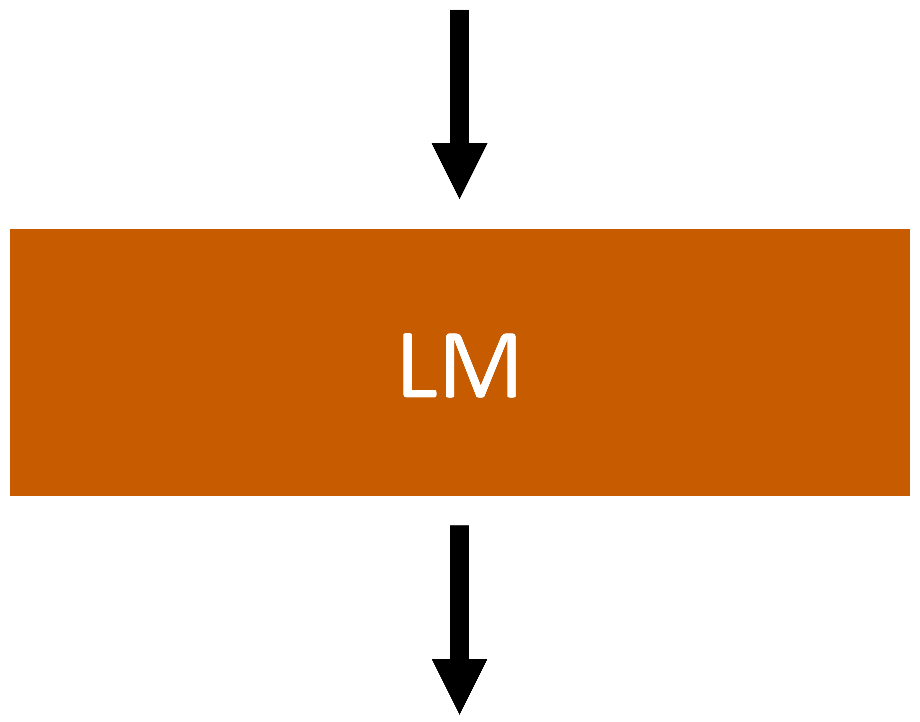
ICL: Advantages

- enables rapid prototyping
- provides a fully natural language interface
- reuses the same model for each task
  - reduces memory requirements and system complexity when serving many different tasks.
- Finetuning can be unstable in the few-shot setting (Schick & Schutze, 2021)

Zhao, Calibrate Before Use: Improving Few-Shot Performance of Language Models, ICML 2021

We don’t know how models in-context learn

	Circulation revenue has increased by 5% in Finland.	Positive
	Panostaja did not disclose the purchase price.	Neutral
	Paying off the national debt will be extremely painful.	Negative
Input Text	The acquisition will have an immediate positive impact.	??



Prediction: Positive

Note

Learns to do a downstream task by conditioning on input-output **examples**

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inferenc, ICLR 2022

We don’t know how models in-context learn

Circulation revenue has increased by 5% in Finland.	Positive
Panostaja did not disclose the purchase price.	Neutral
Paying off the national debt will be extremely painful.	Negative
The acquisition will have an immediate positive impact.	??

Input Text

LM

Prediction: Positive

Note

No weight update and model is not explicitly pre-trained to learn from examples

Question

How does it know what to do then?

- Circulation revenue has increased by 5% in Finland. Positive
- Panostaja did not disclose the purchase price. Neutral
- Paying off the national debt will be extremely painful. Negative
- The company anticipated its operating profit to improve. ??

- Circulation revenue has increased by 5% in Finland. Finance
- Panostaja did not disclose the purchase price. Sports
- Paying off the national debt will be extremely painful. Tech
- The company anticipated its operating profit to improve. ??

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Understanding in-context learning

- A mathematical framework (Xie et al., 2022)
  - **Bayesian inference** view: understand how in-context learning emerges

- Empirical evidence (Hendel et al., 2023)
  - ICL creates task vectors

- Empirical evidence (Min et al., 2022)
  - Which aspects of the prompt affect downstream task performance?

- Circulation revenue has increased by 5% in Finland. Positive
- Panostaja did not disclose the purchase price. Neutral
- Paying off the national debt will be extremely painful. Negative
- The company anticipated its operating profit to improve. ??

LM

Prediction: Positive

- Circulation revenue has increased by 5% in Finland. Finance
- Panostaja did not disclose the purchase price. Sports
- Paying off the national debt will be extremely painful. Tech
- The company anticipated its operating profit to improve. ??

LM

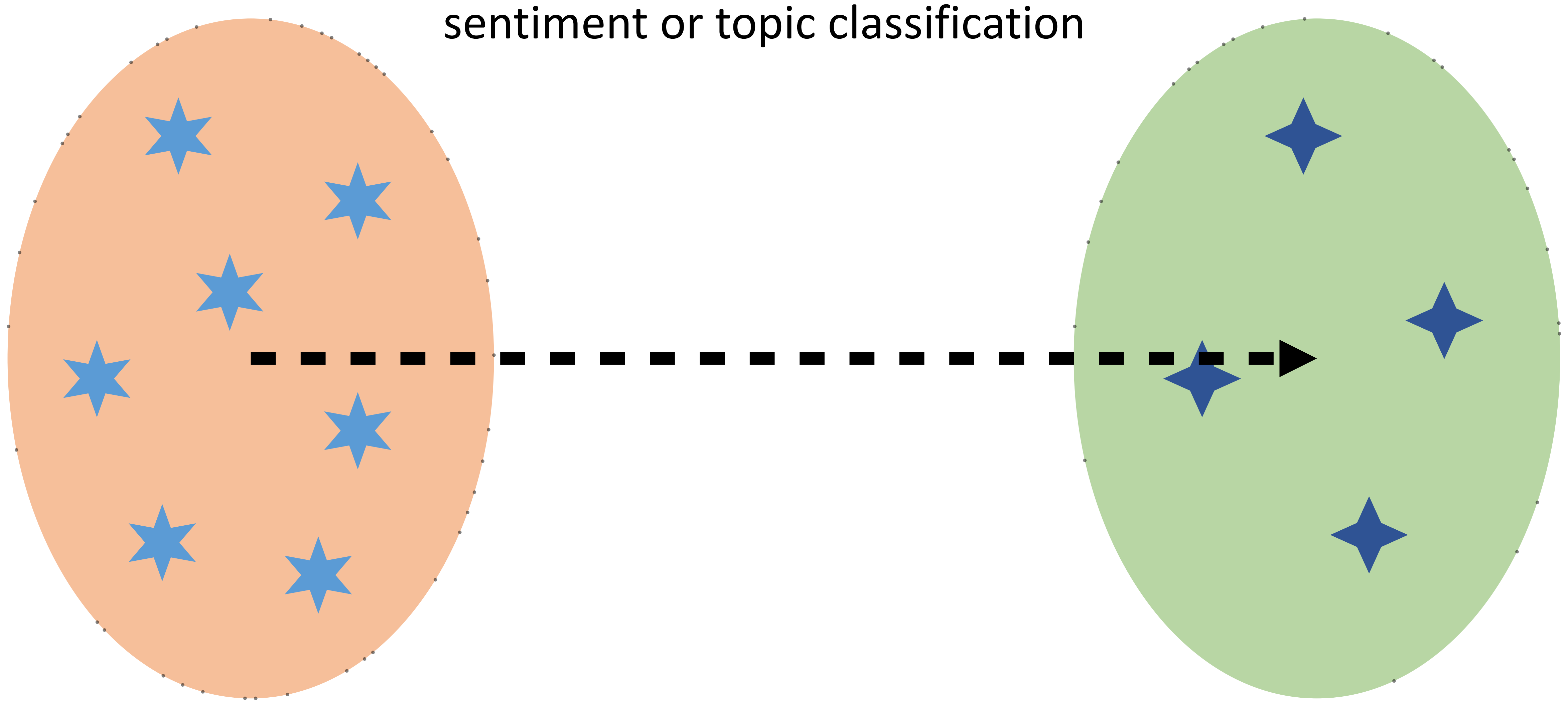
Prediction: Finance

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Model Requirements

Input-Output mapping

sentiment or topic classification



Input distribution

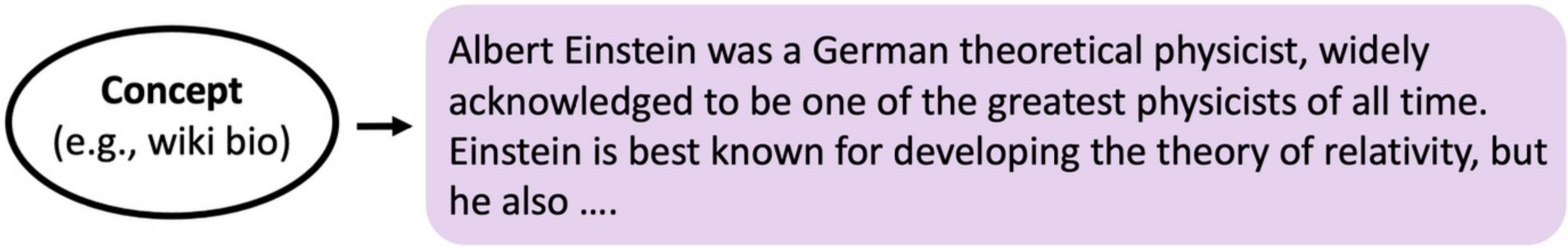
financial or general news

Output distribution

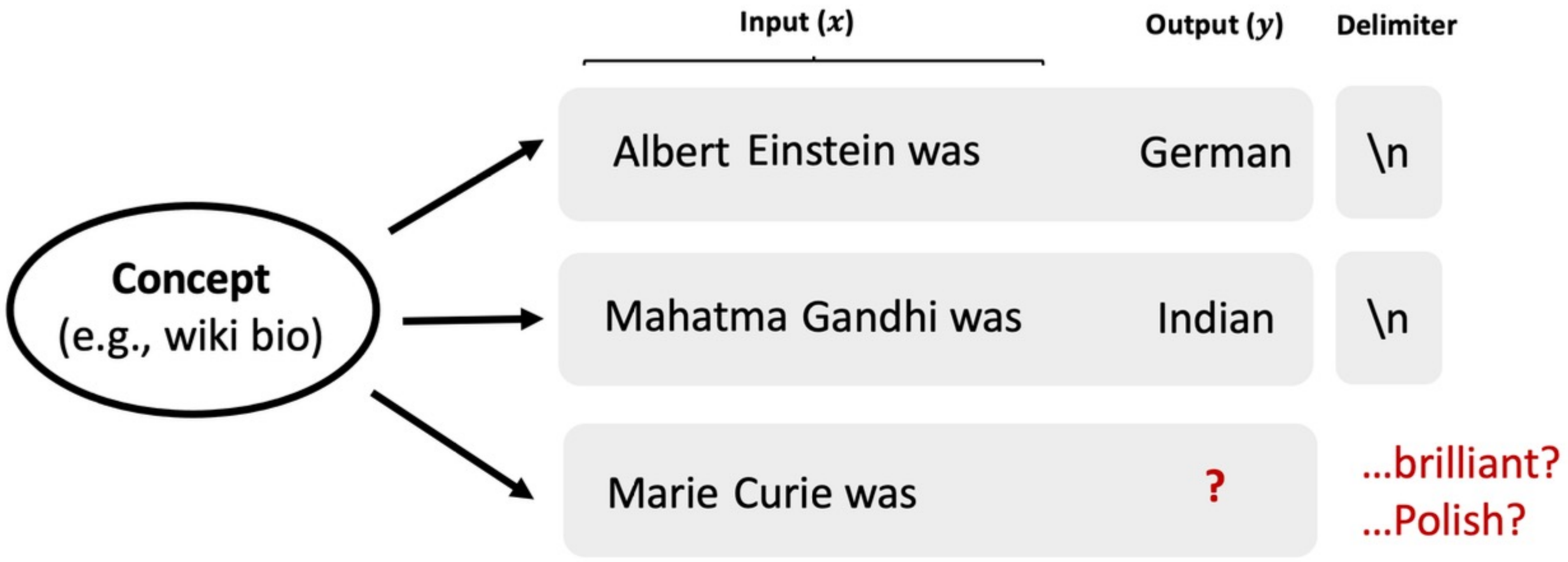
Positive/Negative or topic

Formating

1. Pretraining documents are conditioned on a **latent concept** (e.g., biographical text)



2. Create **independent examples** from a **shared concept.** If we focus on full names, wiki bios tend to relate them to nationalities.



3. Concatenate examples into a prompt and predict next word(s). **Language model (LM) implicitly infers the shared concept** across examples despite the unnatural concatenation

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Concepts (long-term coherence)

- Circulation revenue has increased by 5% in Finland. Positive
- Panostaja did not disclose the purchase price. Neutral
- Paying off the national debt will be extremely painful. Negative
- The company anticipated its operating profit to improve. ??

LM

Infer & retrieve concept

Finance news sentiment

Prediction: Positive

- Circulation revenue has increased by 5% in Finland. Finance
- Panostaja did not disclose the purchase price. Sports
- Paying off the national debt will be extremely painful. Tech
- The company anticipated its operating profit to improve. ??

LM

Infer & retrieve concept

News topics

Prediction: Finance

Note

A latent variable that contains various **document-level statistics**: a distribution of words, a format, a relation between sentences, and other semantic and syntactic relations in general.

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Hypothesis

Language model (LM) uses the in-context learning prompt to "locate" a previously learned concept to do the in-context learning task

Bayesian inference!

$$p(output|prompt) = \int_{concept} p(output|concept, prompt)p(concept|prompt)d(concept)$$

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

How does the LM learn to do Bayesian inference?

- **Pre-train**: To predict the next token during pre-training, the LM must infer the latent concept for the document using evidence from the previous sentences.

- **In-context learning**: If the LM also infers the prompt concept using demonstrations in the prompt, then in-context learning succeeds!

Mixture of Hidden Markov Models (HMM)

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Pre-training distribution

- Each pre-training document is a length T sequence sampled by

$$p(o_1, \ldots, o_T) = \int_{\theta \in \Theta} p(o_1, \ldots, o_T | \theta) p(\theta) d\theta$$

  where θ is a family of concepts that defines a distribution over observed tokens o

- Assumption: $p(o_1, \ldots, o_T | \theta)$ is defined by a Hidden Markov Model (HMM). The concept θ determines the transition probability matrix of the HMM hidden states $h_1, \ldots, h_T$ from a hidden state set H.

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Pre-training distribution

- If the pre-training data is a mixed of finance news **sentiment** task and news **topics** task, intuitively, we could say there are two concepts θ_1 and θ_2.

- p(Paying off the national debt will be extremely painful) =
  ½ p(Paying off the national debt will be extremely painful | θ_1)
  +
  ½ p(Paying off the national debt will be extremely painful | θ_2)

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Prompt distribution

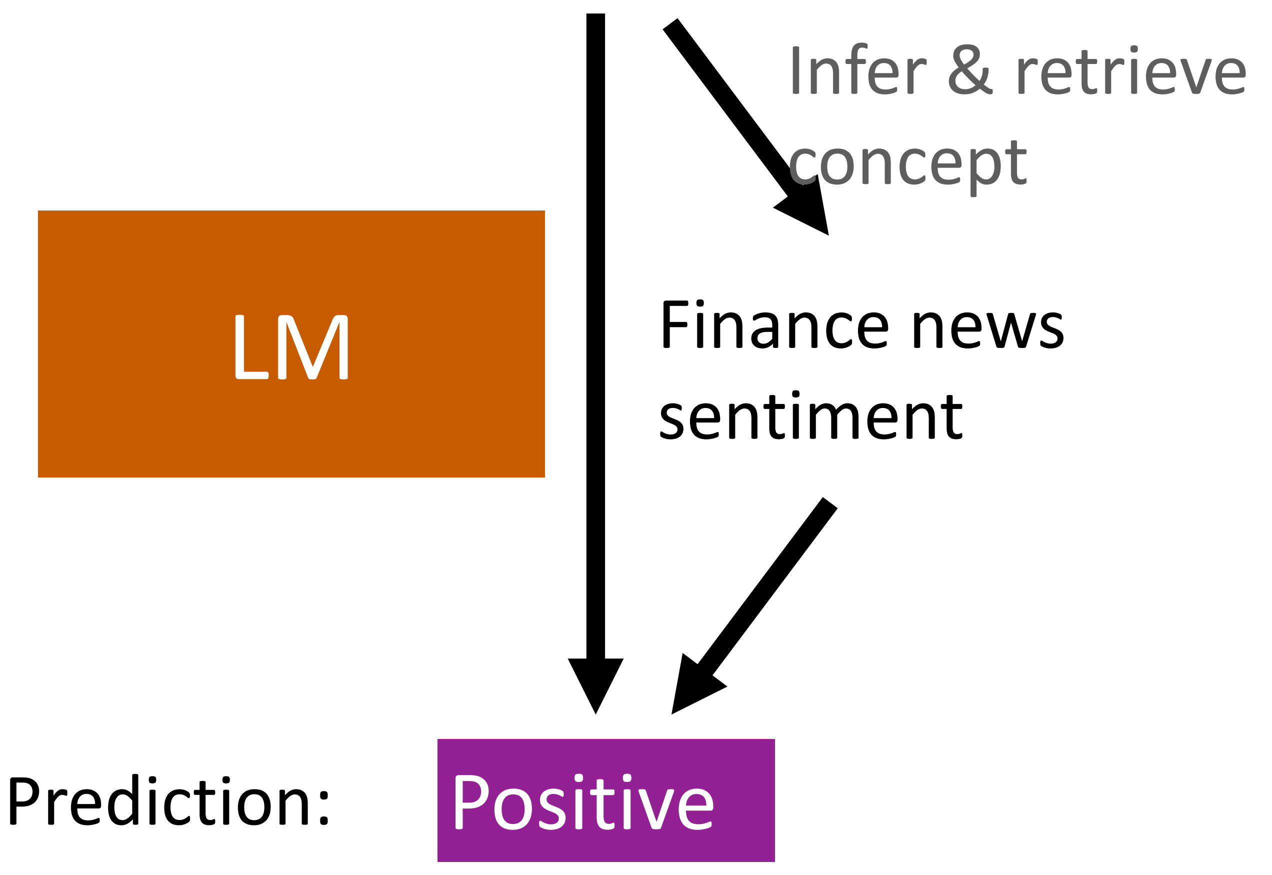
- The prompt is a sequence of demonstrations S_n followed by the test example x_{test} :

$$[S_n, x_{test}] = [x_1, y_1, o^{delim}, x_2, y_2, o^{delim}, \dots, x_n, y_n, o^{delim}, x_{test}]$$

- Given $\theta^* \in \Theta$, the prompt is a concatenation of n independent demonstrations and 1 test input x_{test} that are all conditioned on θ^*.

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

- Circulation revenue has increased by 5% in Finland. Positive
- Panostaja did not disclose the purchase price. Neutral
- Paying off the national debt will be extremely painful. Negative
- The company anticipated its operating profit to improve. ??



[Circulation revenue has increased by 5% in Finland., Positive,
#,
Panostaja did not disclose the purchase price., Neutral,
#,
Paying off the national debt will be extremely painful., Negative,
#,
The company anticipated its operating profit to improve.]

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Under some assumptions, as $n \to \infty$,

$$argmax_y p(y|S_n, X_{test}) \to argmax_y p_{prompt}(y|X_{test})$$

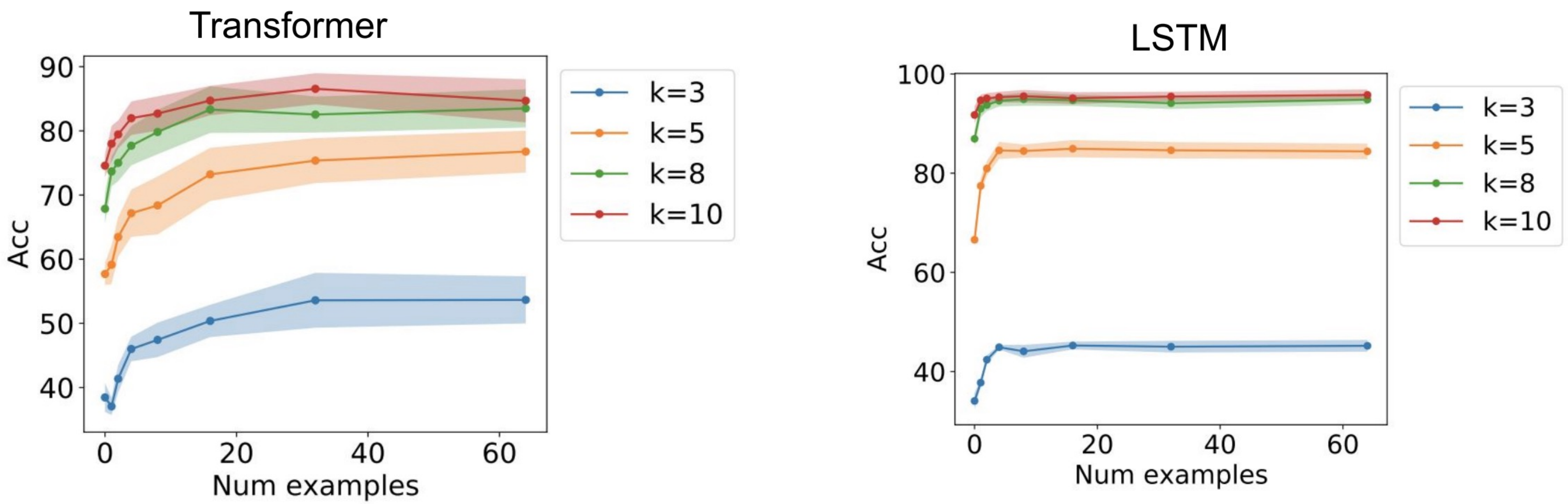
- $p_{prompt} \sim p(.|\theta^*)$
- The in-context predictor asymptotically achieves the optimal expected error
- More examples ⟶ More signals for Bayesian inference ⟶ Smaller error

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

GINC: Generative In-Context learning Dataset

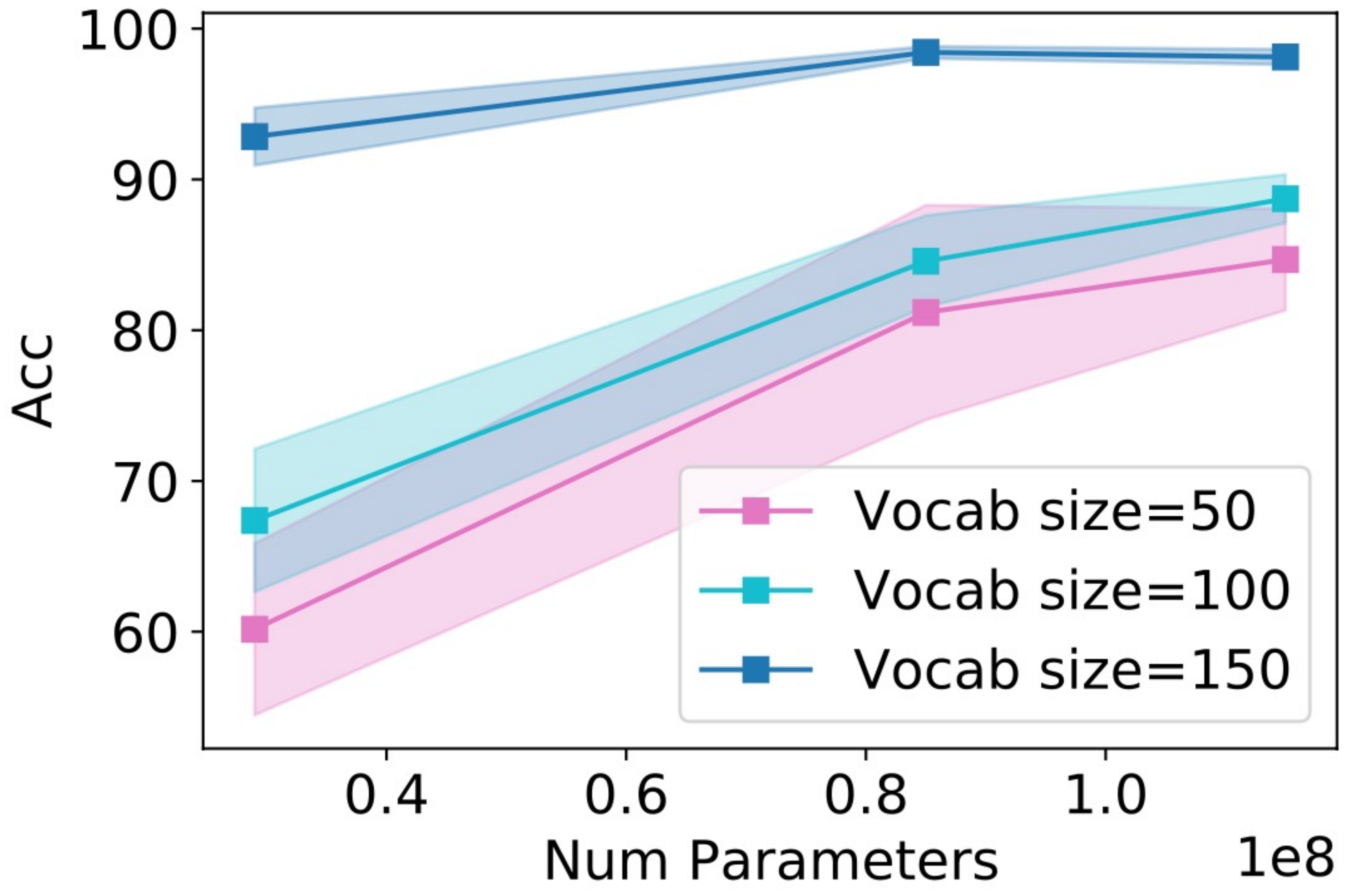
- A synthetic pretraining dataset and in-context learning testbed with the latent concept structure.
- **Pre-training**: a uniform mixture of HMMs over a family of 5 concepts, 1000 pre-training documents, ~10 million tokens in total
- **Prompts**: 0~64 training examples, example length k=3, 5, 8, 10
- GPT-2-based Transformers and LSTMs
- Vocabulary size: 50, 100, 150

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022



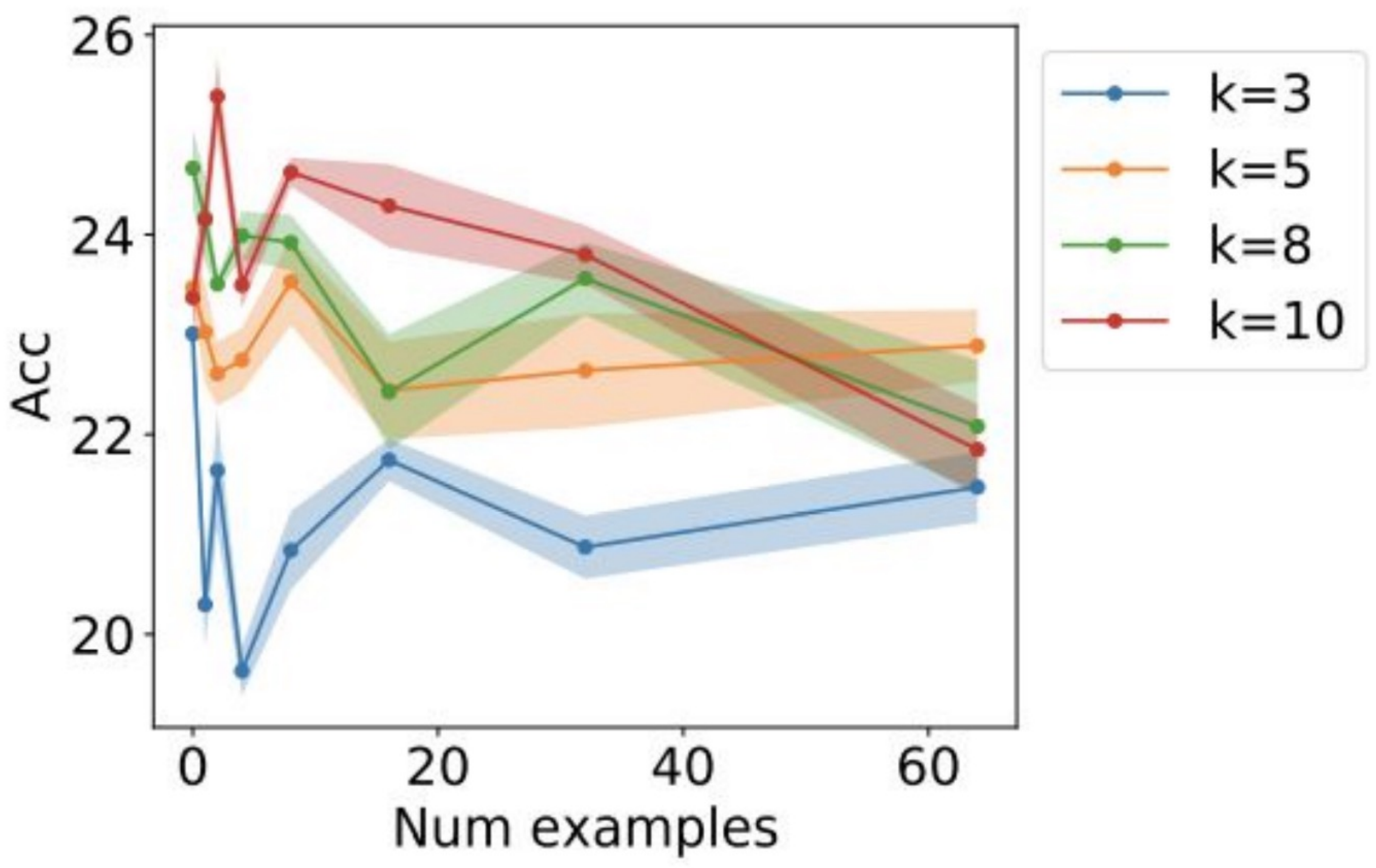
Accuracy increases with number of examples n and length of each example k, which is consistent with the theoretical results.

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

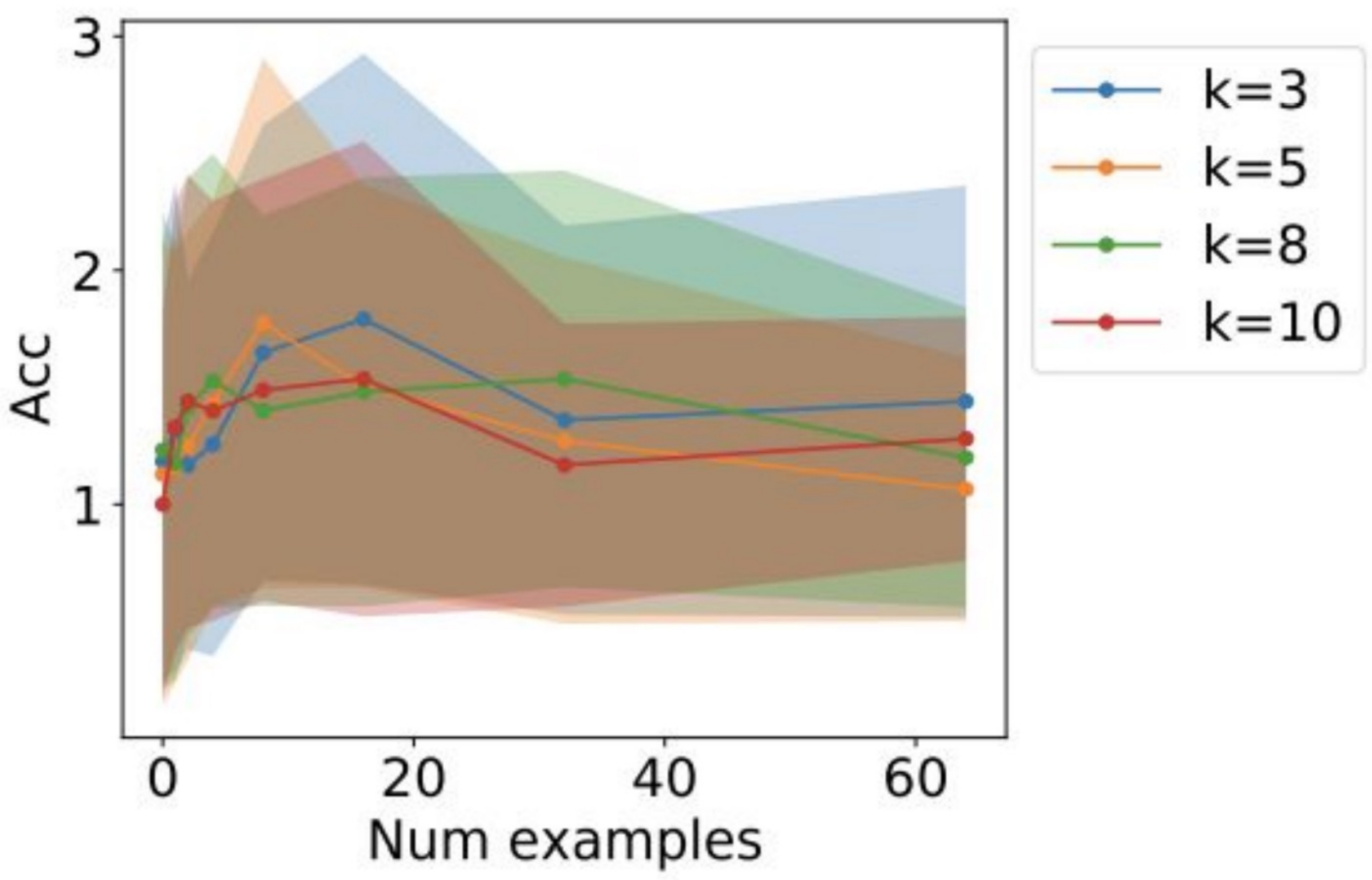


In-context accuracy (95% intervals) of Transformers improves as model size increases on the GINC dataset

Is the HMMs assumption necessary?



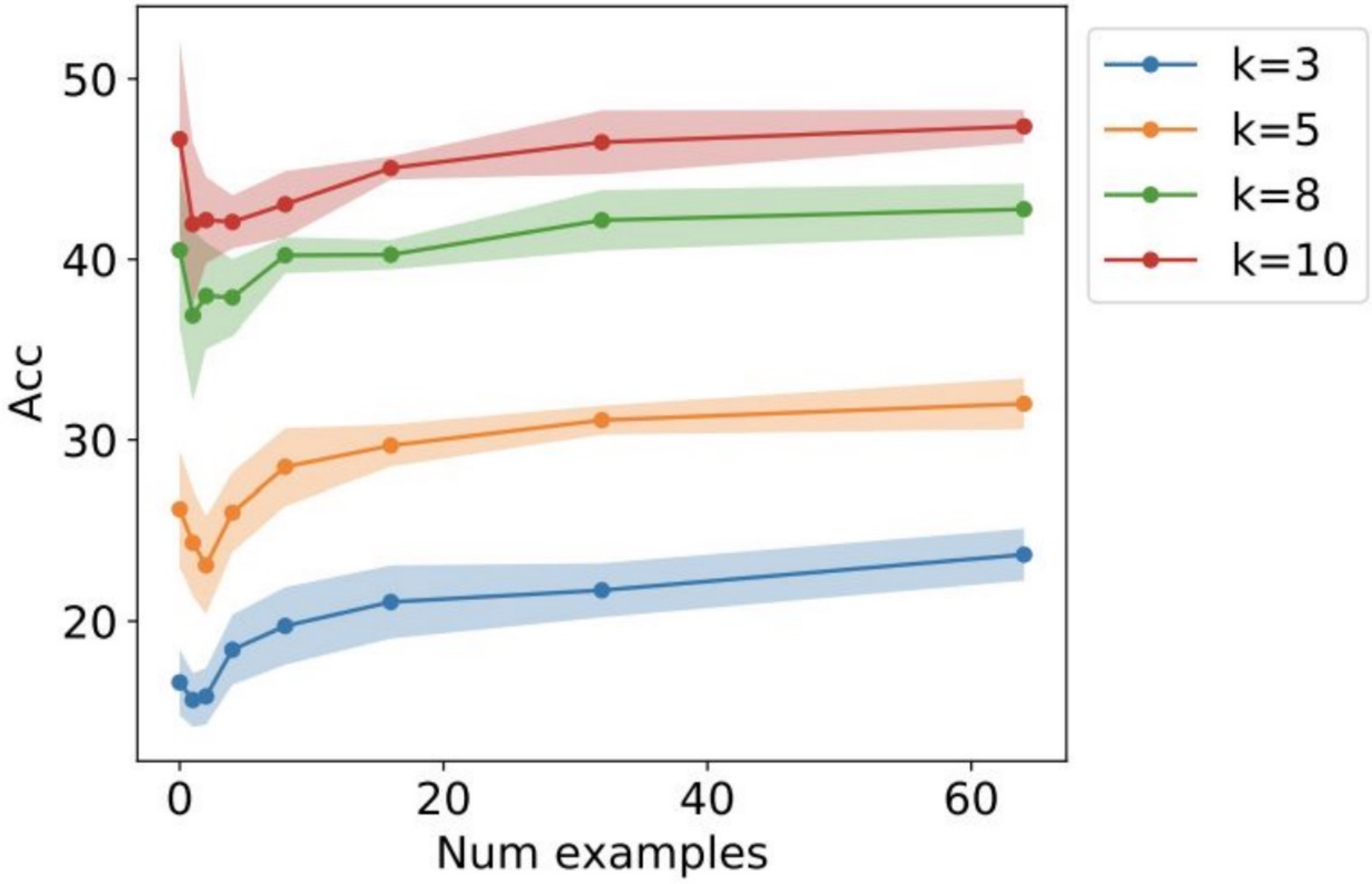
When pre-trained with only one concept, in-context learning fails.



When the pre-training data has random transitions, in-context learning fails.

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

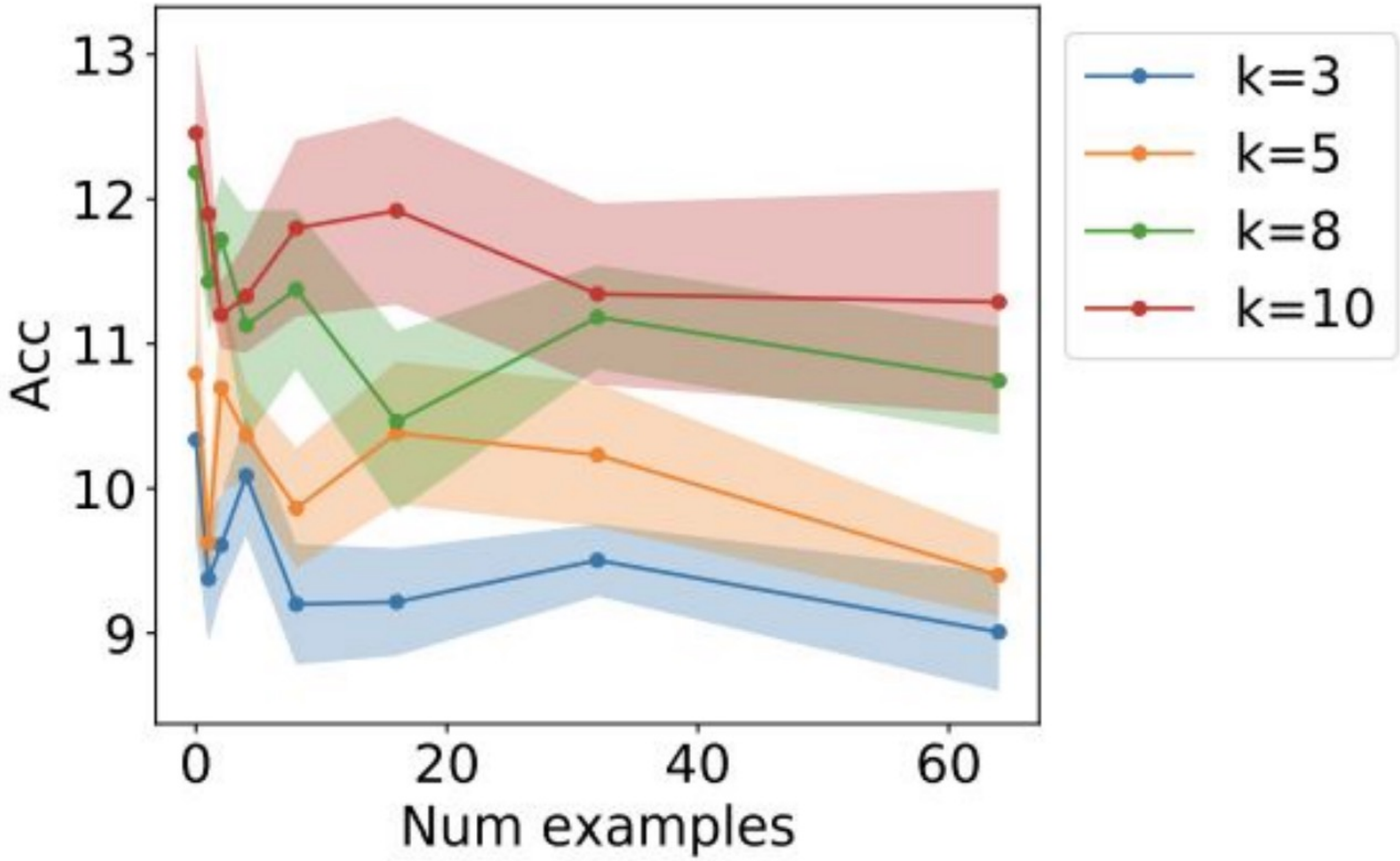
Zero-shot vs One-shot



- In some settings, few-shot accuracy is initially worse than zero-shot accuracy, but can recover with more examples.
- Mirroring the behavior of GPT-3 on some datasets such as LAMBADA, HellaSwag, PhysicalQA, RACE-m, CoQA/SAT
- Especially because the transition probabilities in GINC are lower entropy

Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Unseen Concepts



When prompts are from random unseen concepts, in-context learning fails to extrapolate.

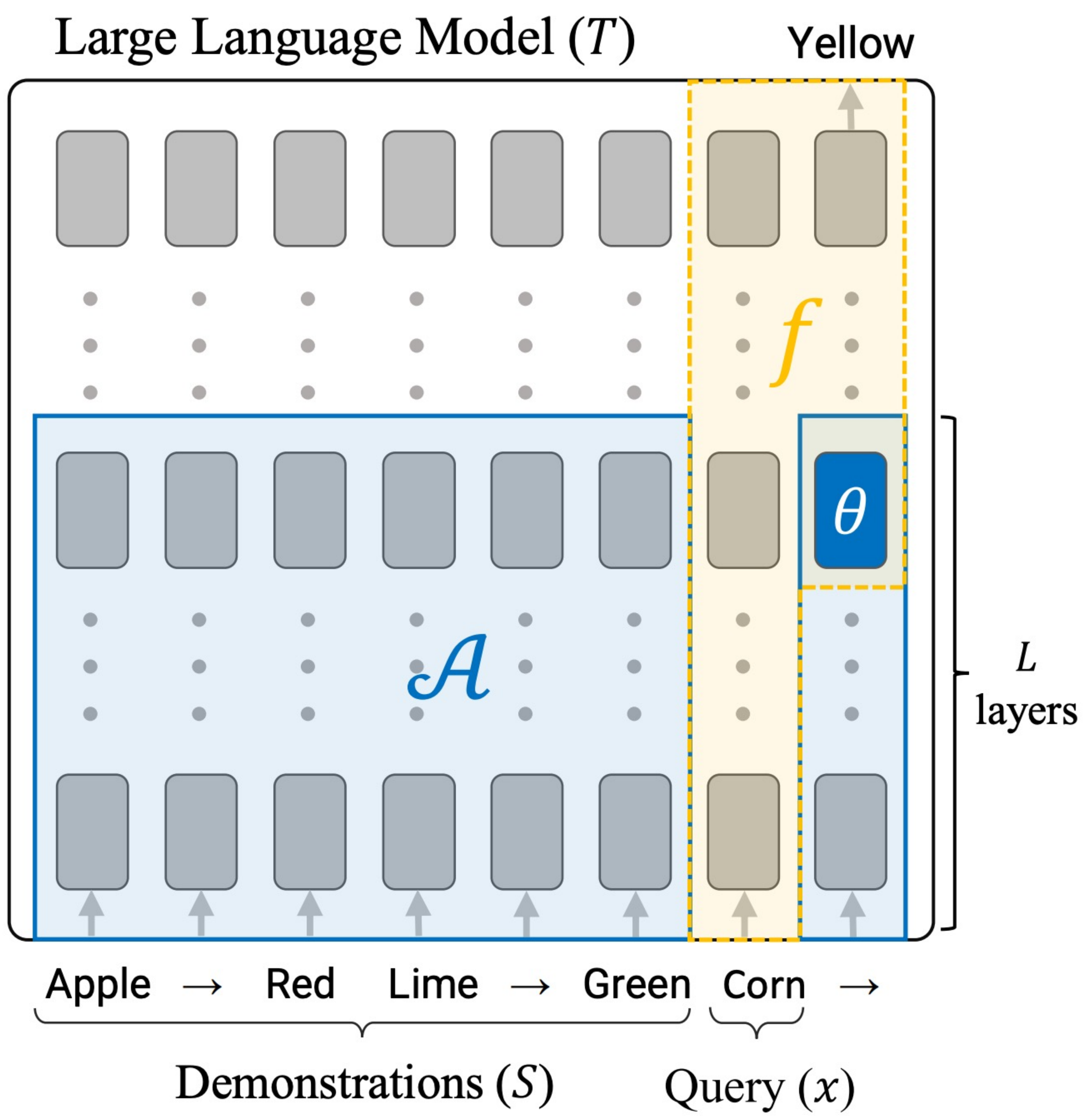
Xie et al., An Explanation of In-context Learning as Implicit Bayesian Inference, 2022

Understanding in-context learning

- A mathematical framework (Xie et al., 2022)
  - **Bayesian inference** view: understand how in-context learning emerges
- Empirical evidence (Hendel et al., 2023)
  - ICL creates task vectors
- Empirical evidence (Min et al., 2022)
  - Which aspects of the prompt affect downstream task performance?

ICL creates task vectors

- A decoder-only transformer T is applied on the demonstrations S and the input x
- $T([S, x]) = f(x;\ A(S))$
  - first maps S into a "task vector" $\theta = A(s)$, independent of x.
  - Then, maps x to the output, based on $\theta \equiv A(S)$, without direct dependence on S.

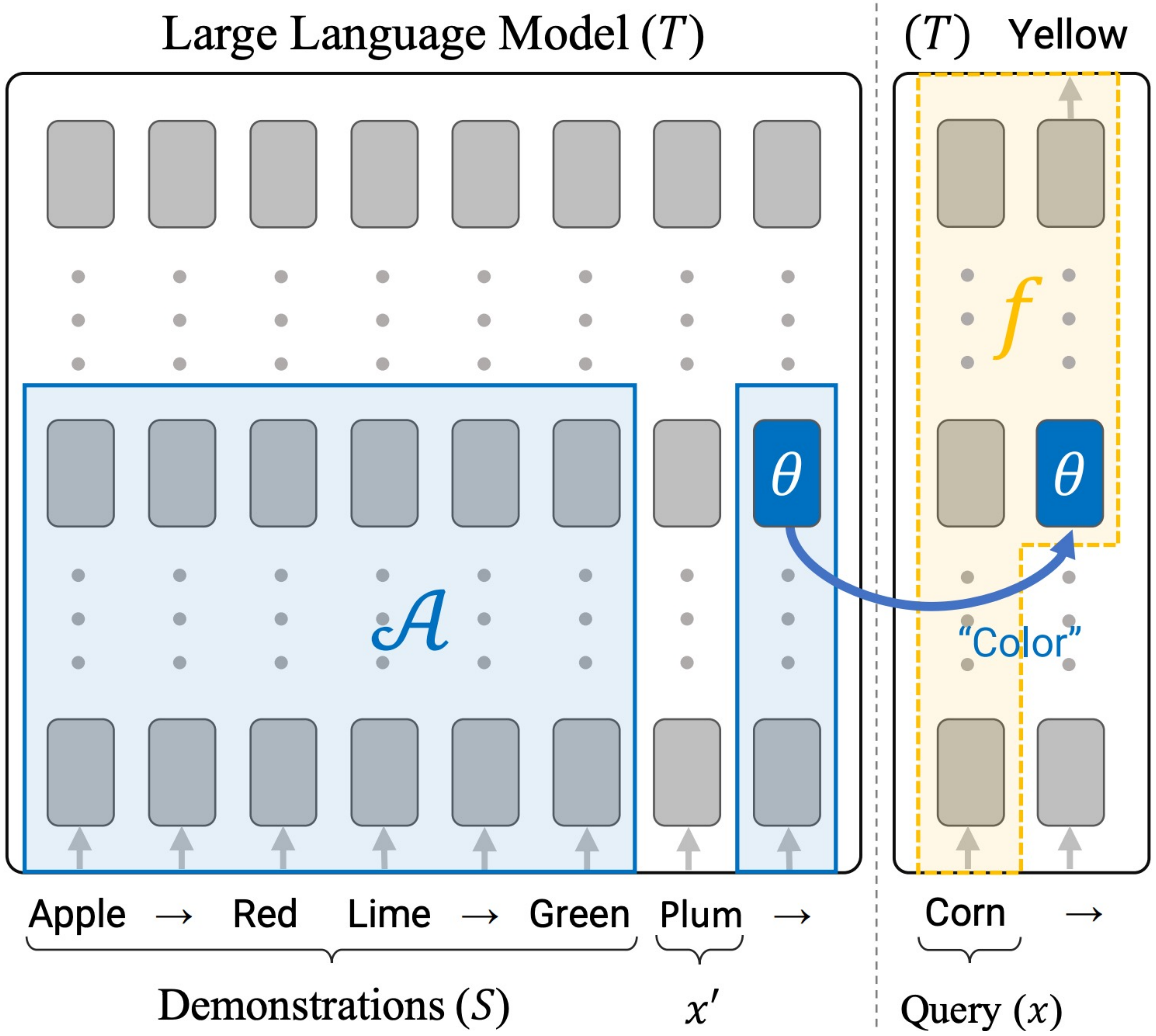


The function f is very similar to the original transformer applied to x without demonstrations but instead modulated by θ

Hendel et al., In-Context Learning Creates Task Vectors, 2023

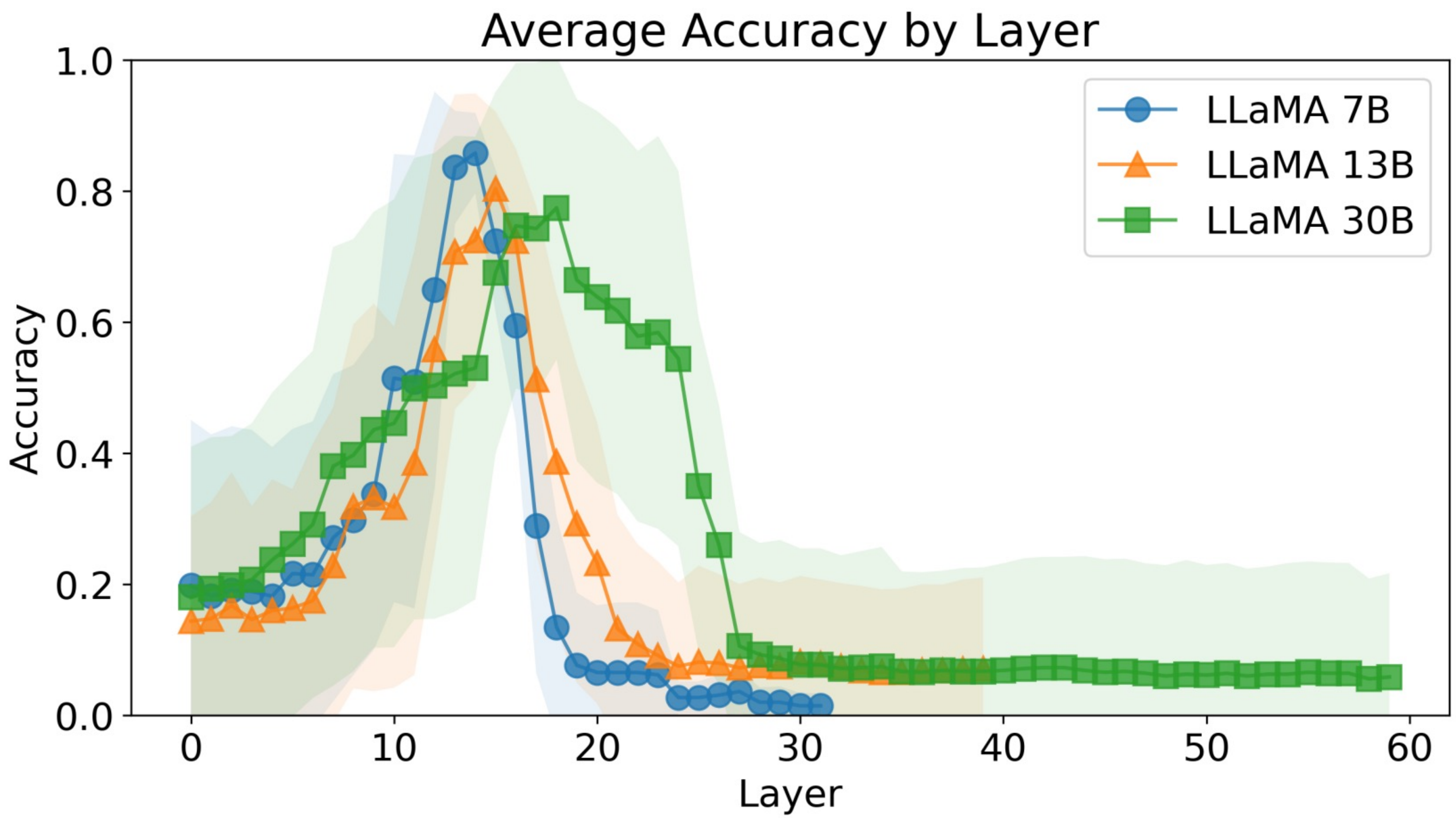
How to separate A and f?

- Consider the layer L where A ends and f begins
- A generates θ using a dummy x'
- $f(\cdot;\ \theta)$ is applied to x by running the transformer on $[x, \rightarrow]$ with θ patched at layer L of $\rightarrow$.



Hendel et al., In-Context Learning Creates Task Vectors, 2023

Which layer to extract task vector?



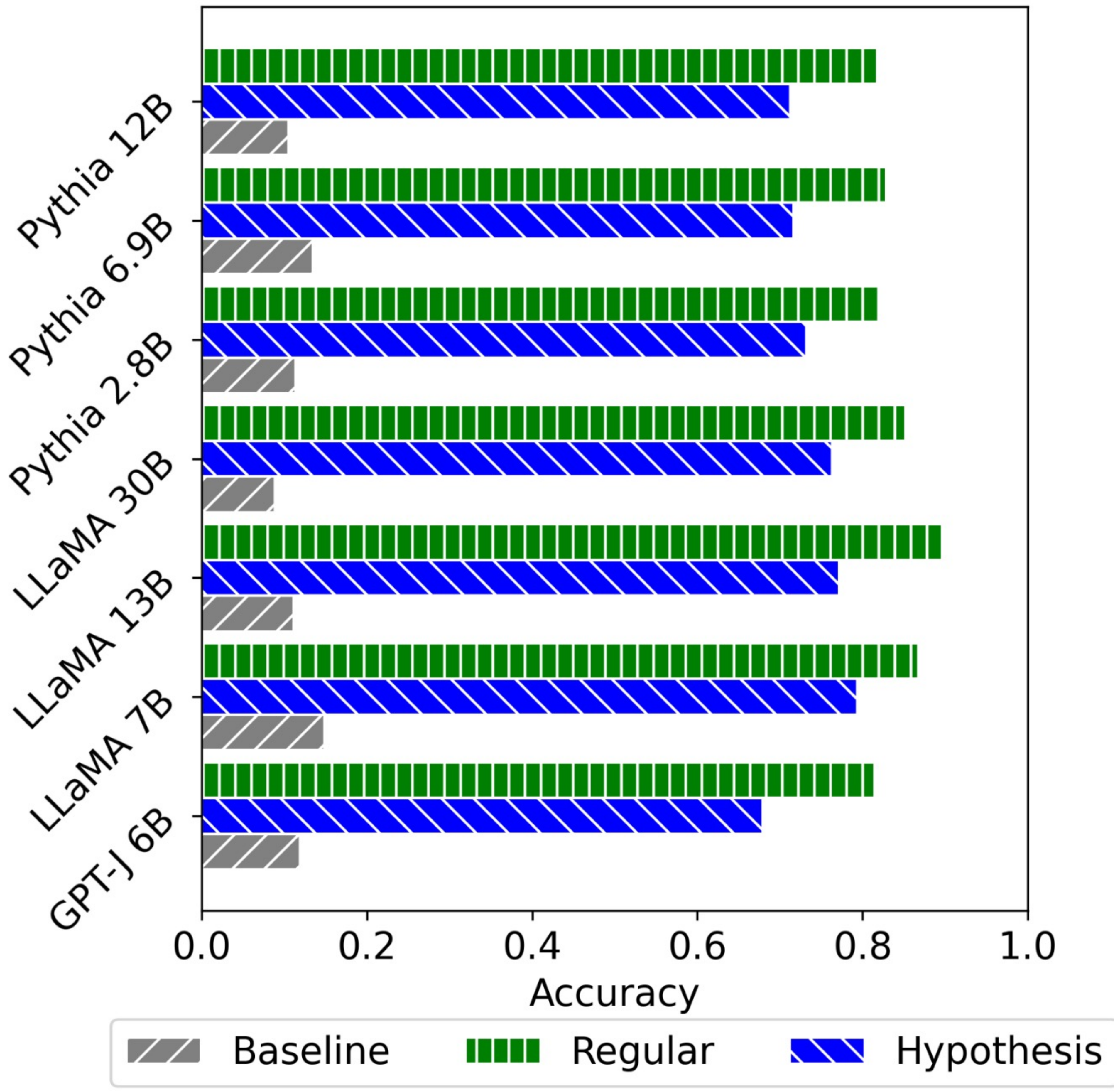
Hendel et al., In-Context Learning Creates Task Vectors, 2023

Separation is not such harmful

- Regular ICL: $T([S, x])$

- Hypothesis: A generates θ using a dummy x', and $f(x;\ \theta)$ is computed by running the transformer on $[x, \rightarrow]$ with θ patched at layer L of $\rightarrow$.

- Baseline: $T([x, \rightarrow])$ without demonstrations S



Hendel et al., In-Context Learning Creates Task Vectors, 2023

Understanding in-context learning

- A mathematical framework (Xie et al., 2022)
  - **Bayesian inference** view: understand how in-context learning emerges
- Empirical evidence (Hendel et al., 2023)
  - ICL creates task vectors
- Empirical evidence (Min et al., 2022)
  - Which aspects of the prompt affect downstream task performance?

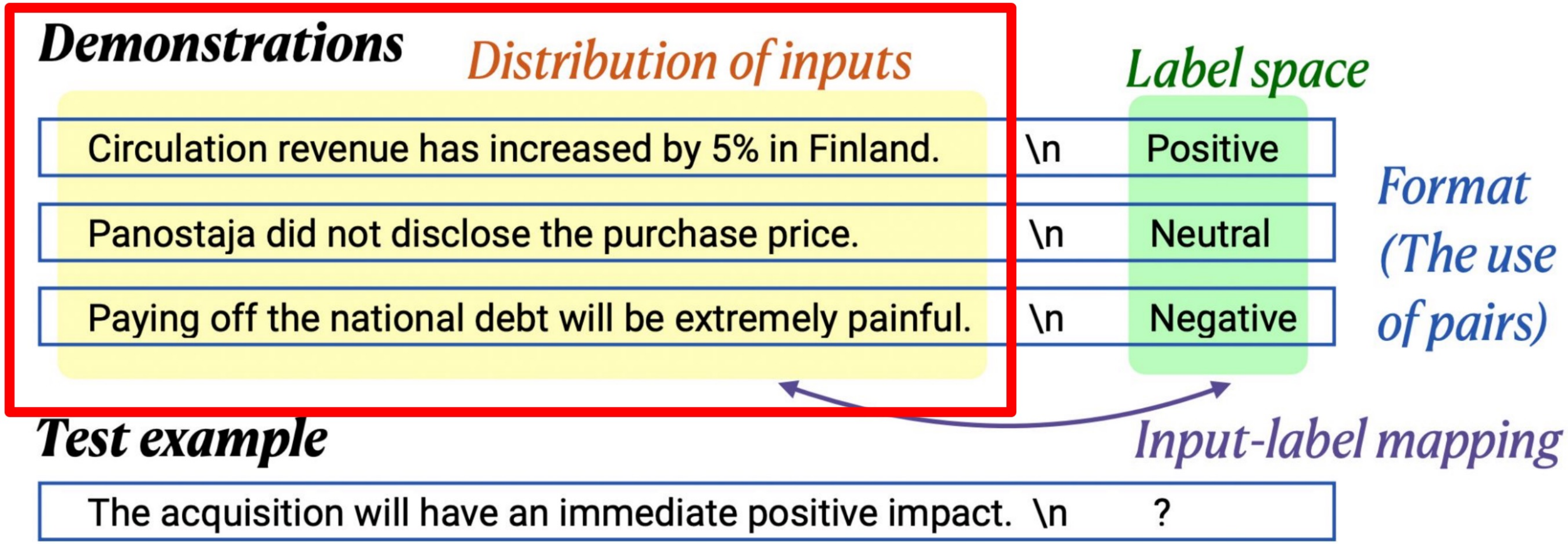
We break the prompt into four parts that provide signal to the model

1 Distribution of Inputs

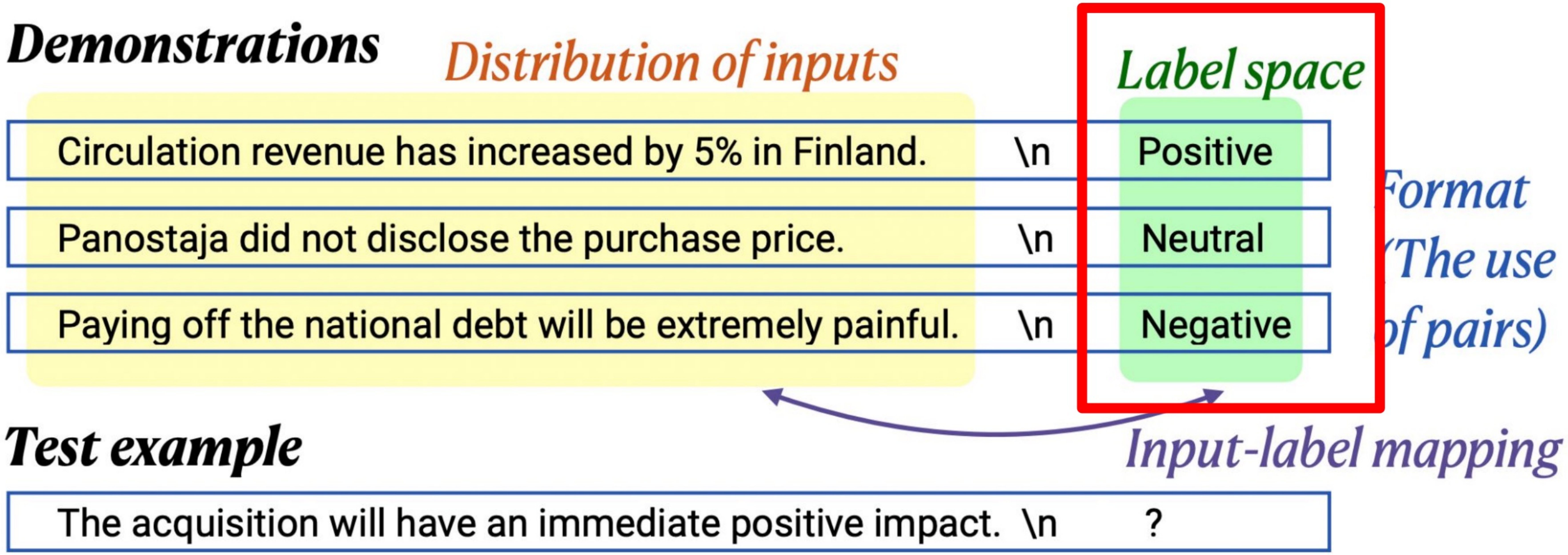
2 Label Space

3 Input-label Mapping

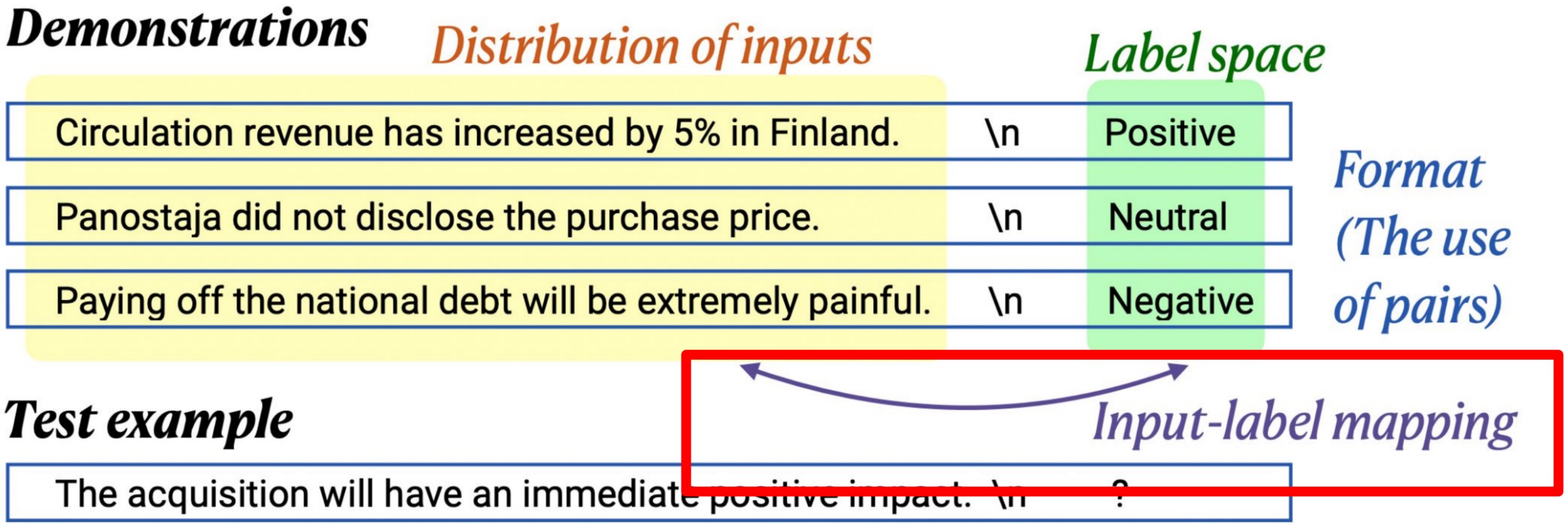
4 Format



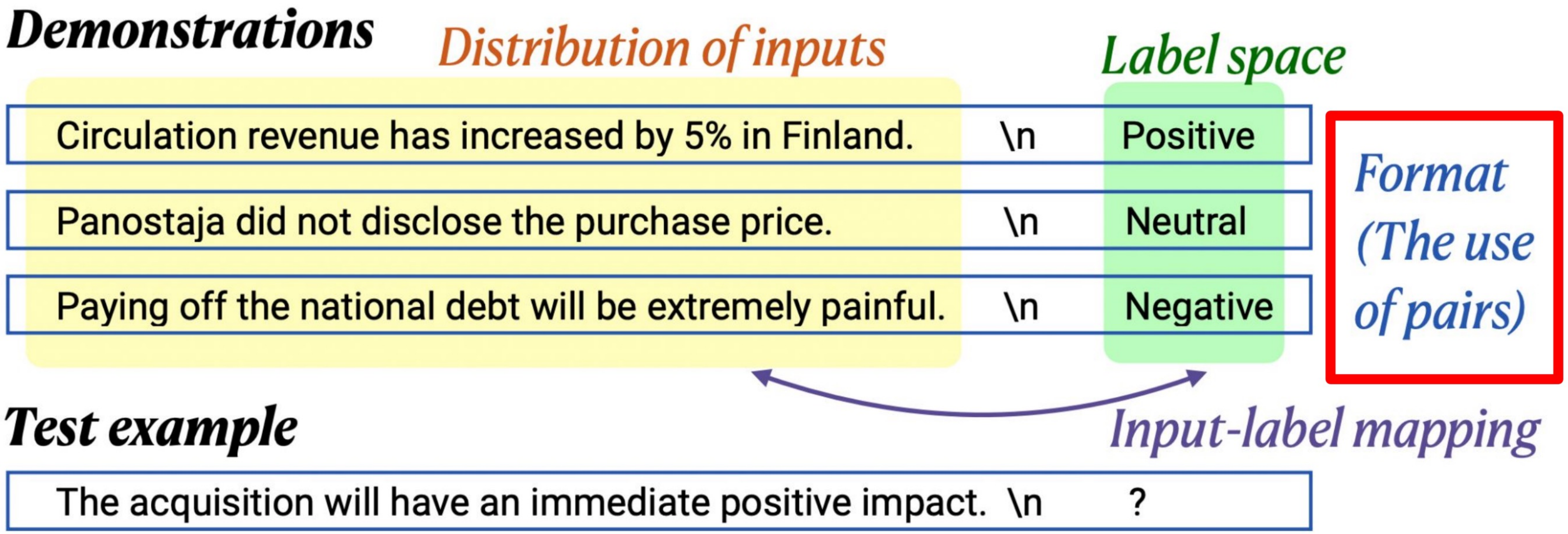
Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022



Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

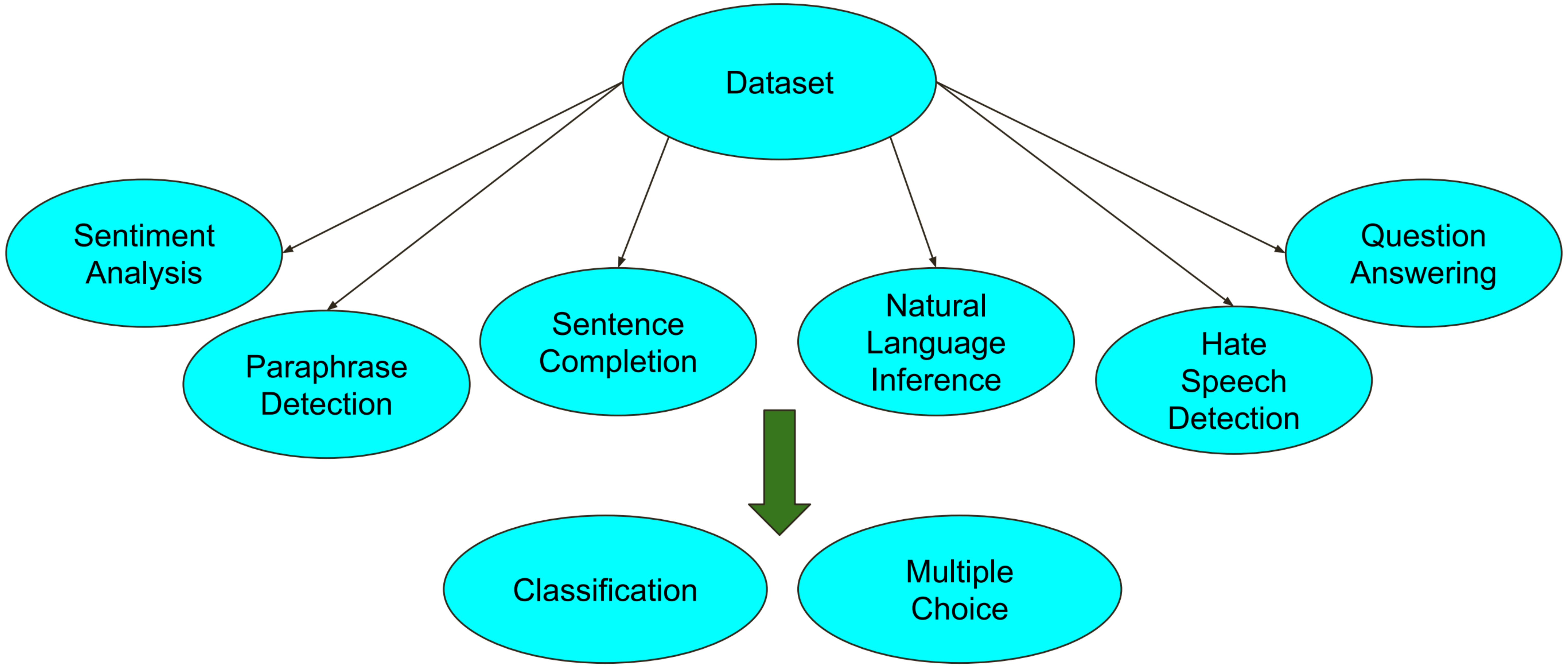


Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022



Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Datasets



Classification Tasks	
Sentiment Analysis	• financial_phrasebank • poem_sentiment
Paraphrase detection	• medical_questions_pairs • glue-mrpc
Natural language inference	• glue-wnli • climate fever
Question Answering	• quarel • openbookqa

Multiple Choice Tasks	
Hate speech detection	• hate_speech18 • ethos-national_origin
Sentence completion	• codah • superglue-copa

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Evaluation Methodology

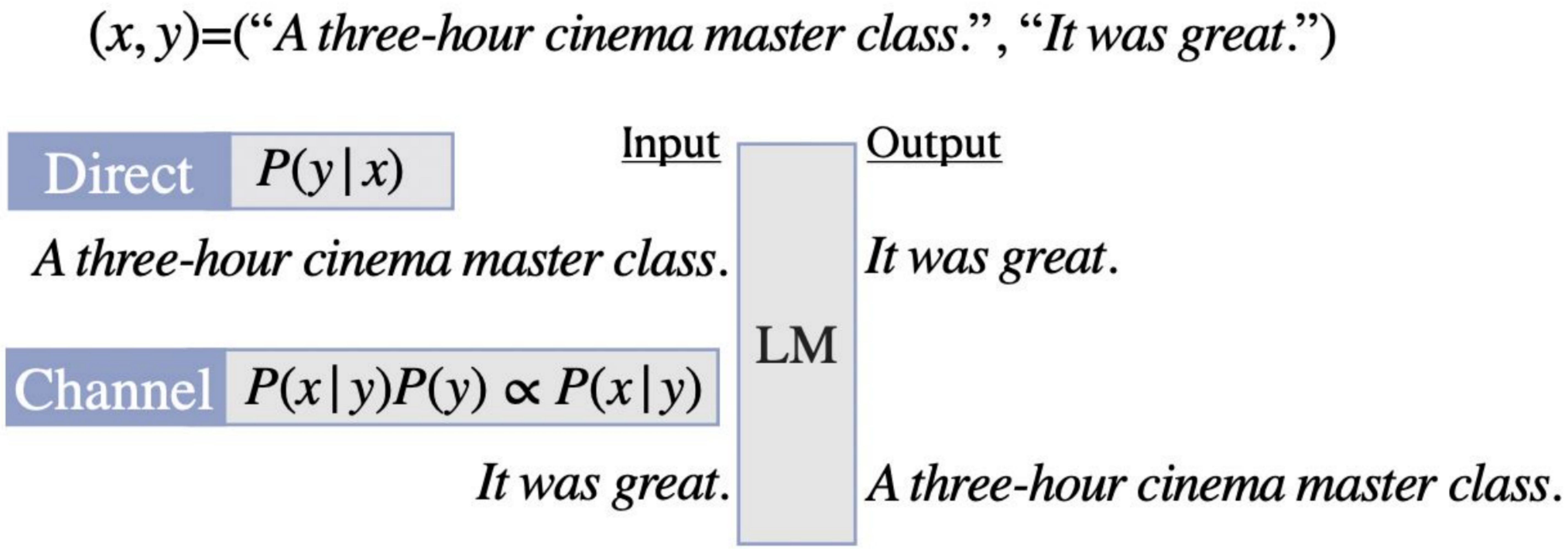
- Metrics
  - Classification: **Macro-F1**
  - Multiple Choice: **Accuracy**
- Compute per-dataset average across seeds, and report **macro-average over datasets**

Models

Model	# Params	Public	Meta-trained
GPT-2 Large	774M	✓	✗
MetaICL	774M	✓	✓
GPT-J	6B	✓	✗
fairseq 6.7B†	6.7B	✓	✗
fairseq 13B†	13B	✓	✗
GPT-3	175B‡	✗	✗

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Direct vs. Channel Models

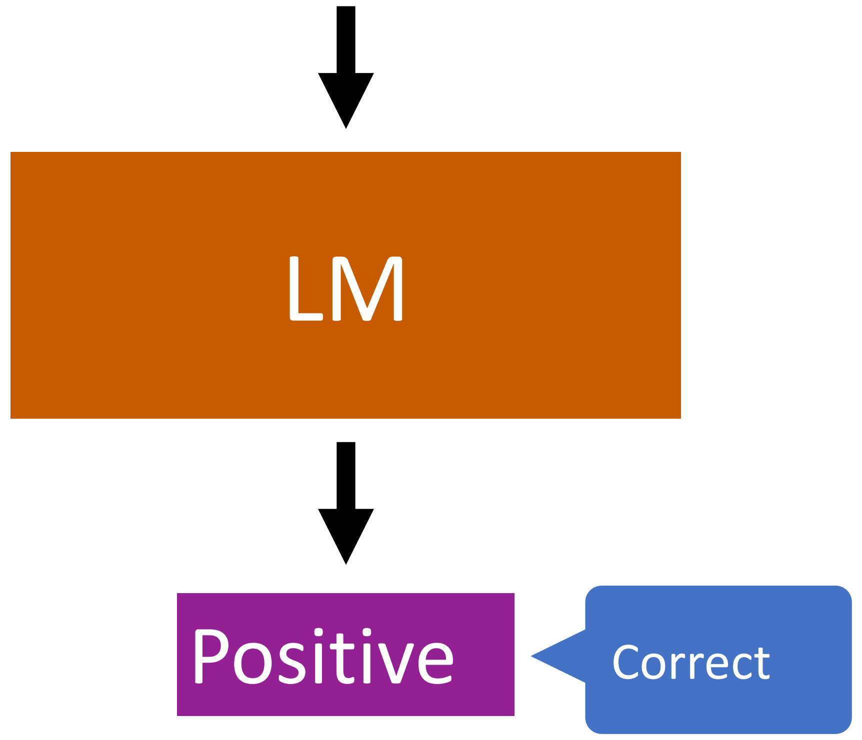


$P(y|x) \longrightarrow P(x|y)$

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

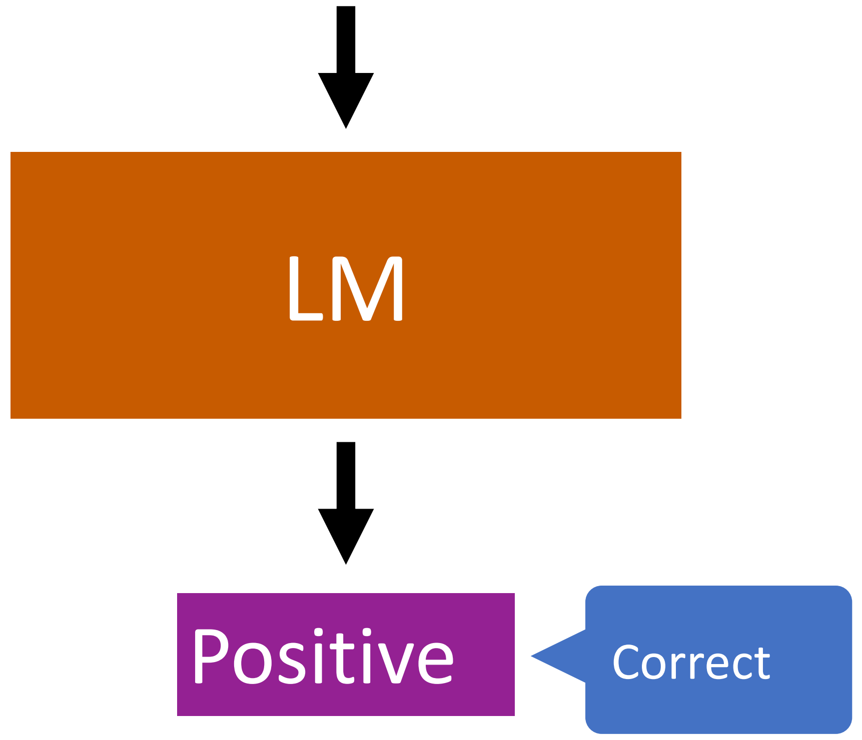
True Labels vs Random Labels

Circulation revenue has increased by 5% in Finland.	Positive
Panostaja did not disclose the purchase price.	Neutral
Paying off the national debt will be extremely painful.	Negative
The acquisition will have an immediate positive impact.	??



Prompt with true labels

Circulation revenue has increased by 5% in Finland.	Neutral
Panostaja did not disclose the purchase price.	Negative
Paying off the national debt will be extremely painful.	Positive
The acquisition will have an immediate positive impact.	??



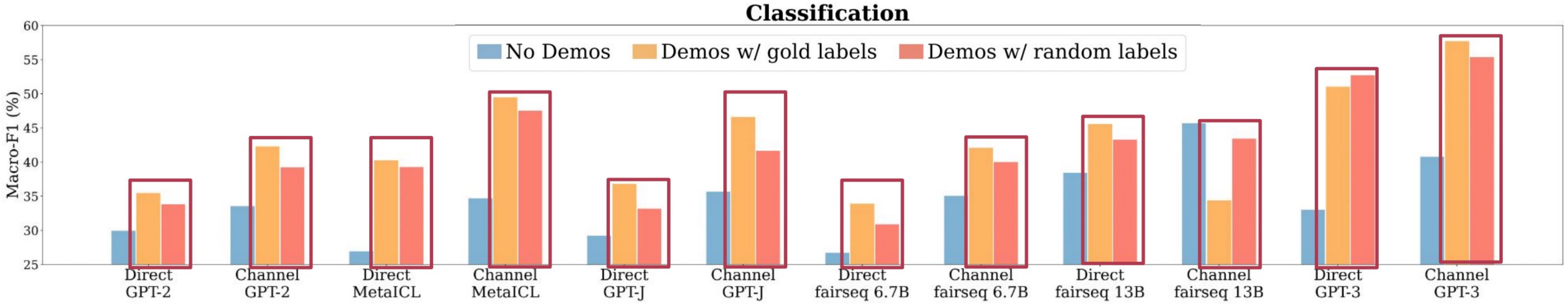
Prompt with random labels

Note

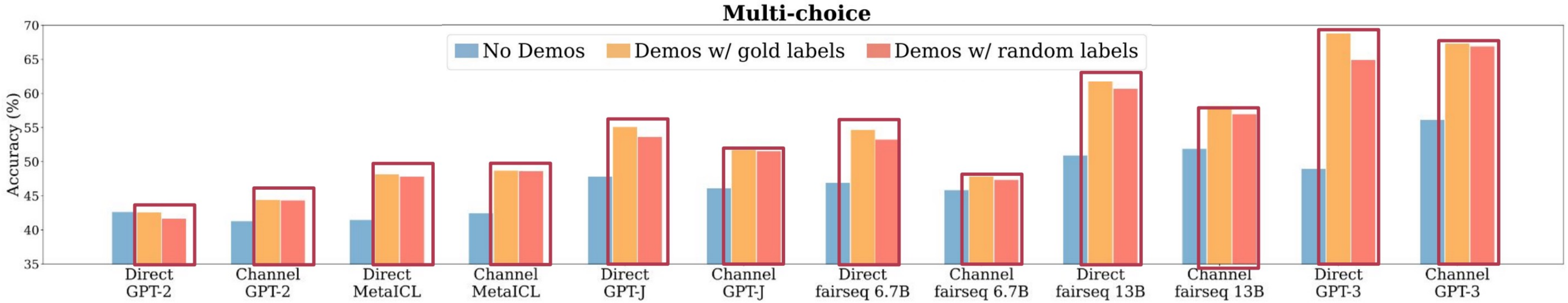
1. Randomly sample a label from the correct label space
2. Assign the label to the example

https://ai.stanford.edu/blog/understanding-incontext/

Results



Comparisons between no-examples (blue), examples with ground truth outputs (yellow) and examples with random outputs (red)



Comparisons between no-examples (blue), examples with ground truth outputs (yellow) and examples with random outputs (red)

Models see small performance drop in the range of 0–5% absolute with random labels

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

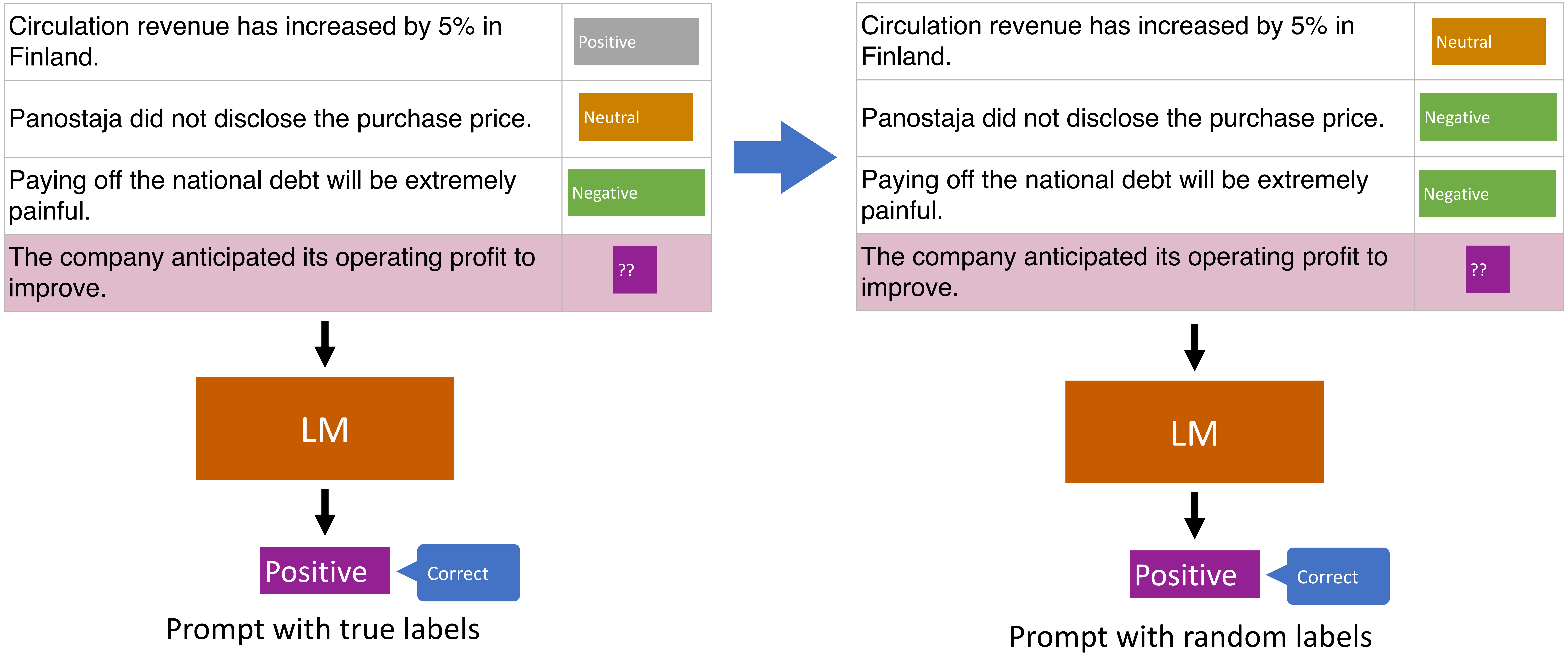
Results Takeaways

- Ground truth input-label mapping in the prompt is not as important as we thought
- Model is not recovering the expected input-label correspondence for the task from the input-label pairings

Question

Is this result consistent in other setups?

Does the number of correct labels matter?

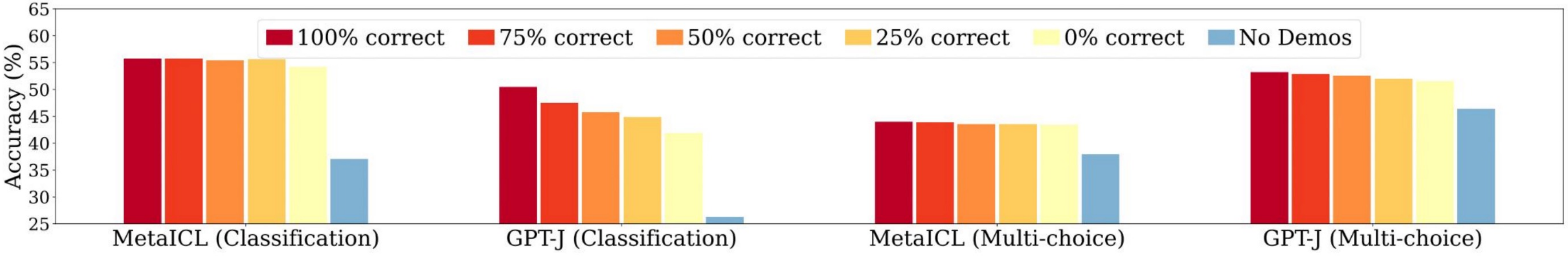


Note

1. Vary the number of correct labels in examples

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Results



Results of varying number of correct labels in the prompt

Model performance is fairly insensitive to the number of correct labels

Using incorrect labels is better than no examples

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Varying the Number of Examples

Circulation revenue has increased by 5% in Finland.	Positive
Panostaja did not disclose the purchase price.	Neutral
Paying off the national debt will be extremely painful.	Negative
The company anticipated its operating profit to improve.	??

LM

Positive — Correct

Prompt with three examples

Circulation revenue has increased by 5% in Finland.	Positive
Panostaja did not disclose the purchase price.	Neutral
The company anticipated its operating profit to improve.	??

LM

Positive — Correct

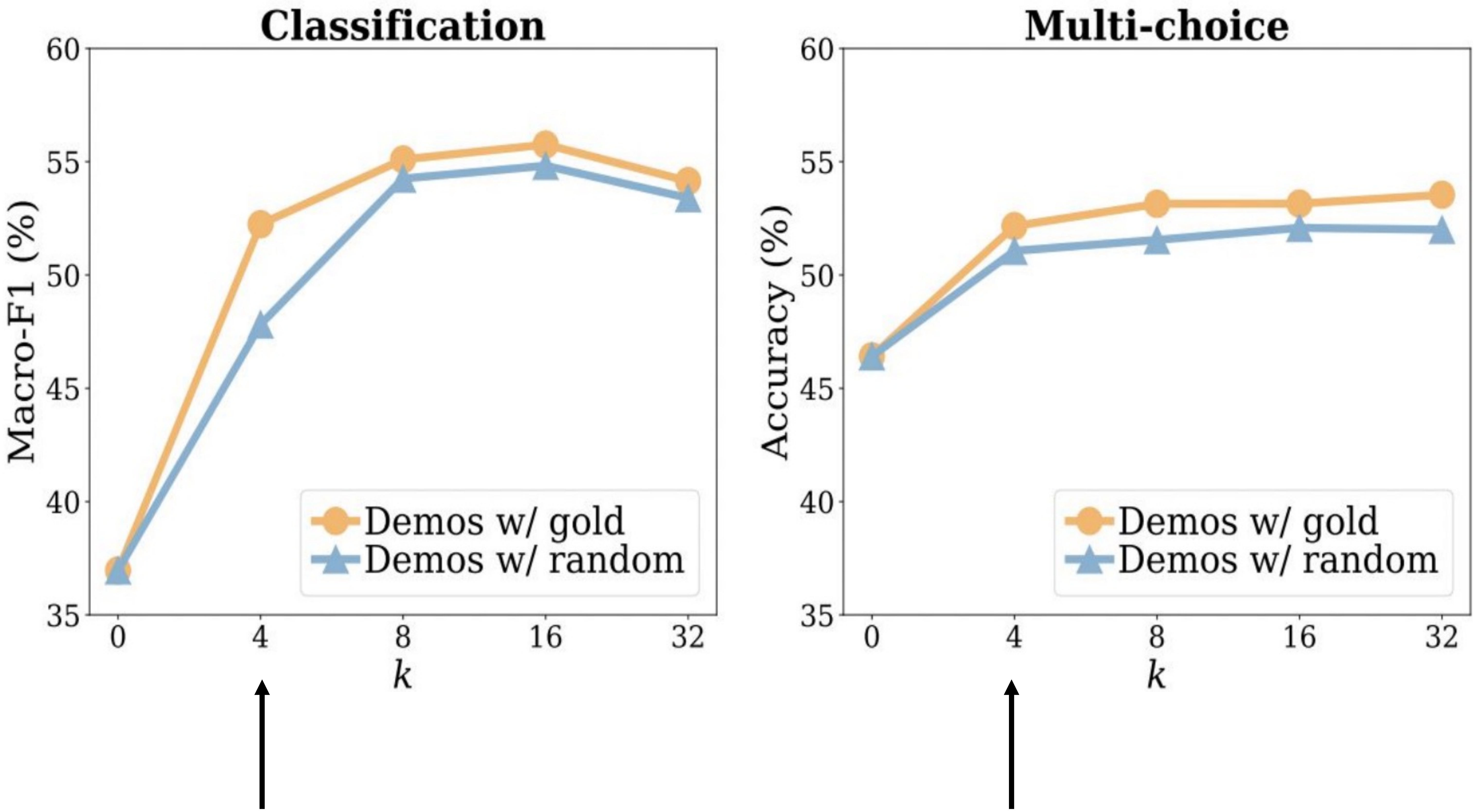
Prompt with two examples

Note

Measure whether the results of using random labels is consistent across differing number of examples

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Results

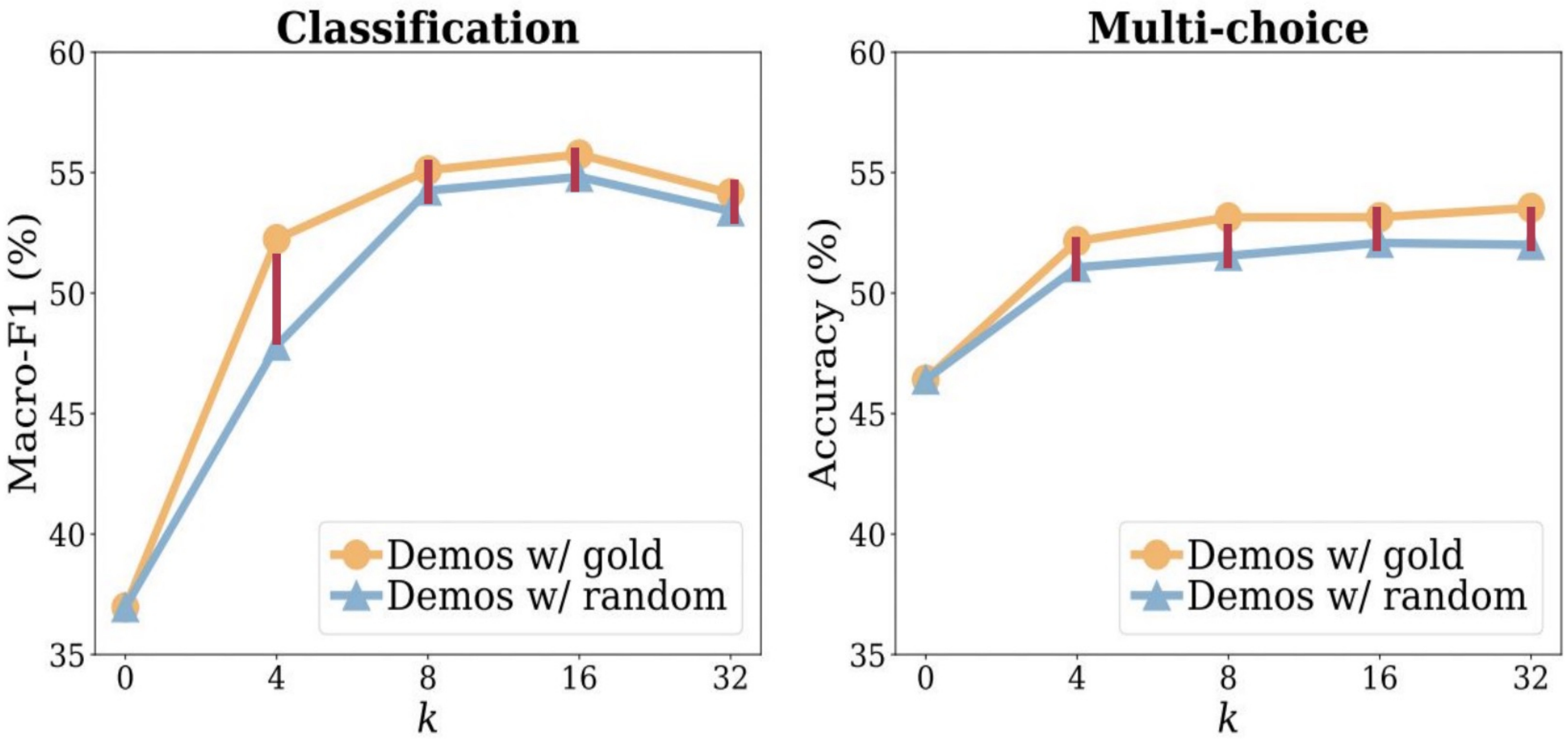


Ablations on varying numbers of examples (k) in the prompt

Using **small number** of examples with **random labels** is better than no **examples**

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Results

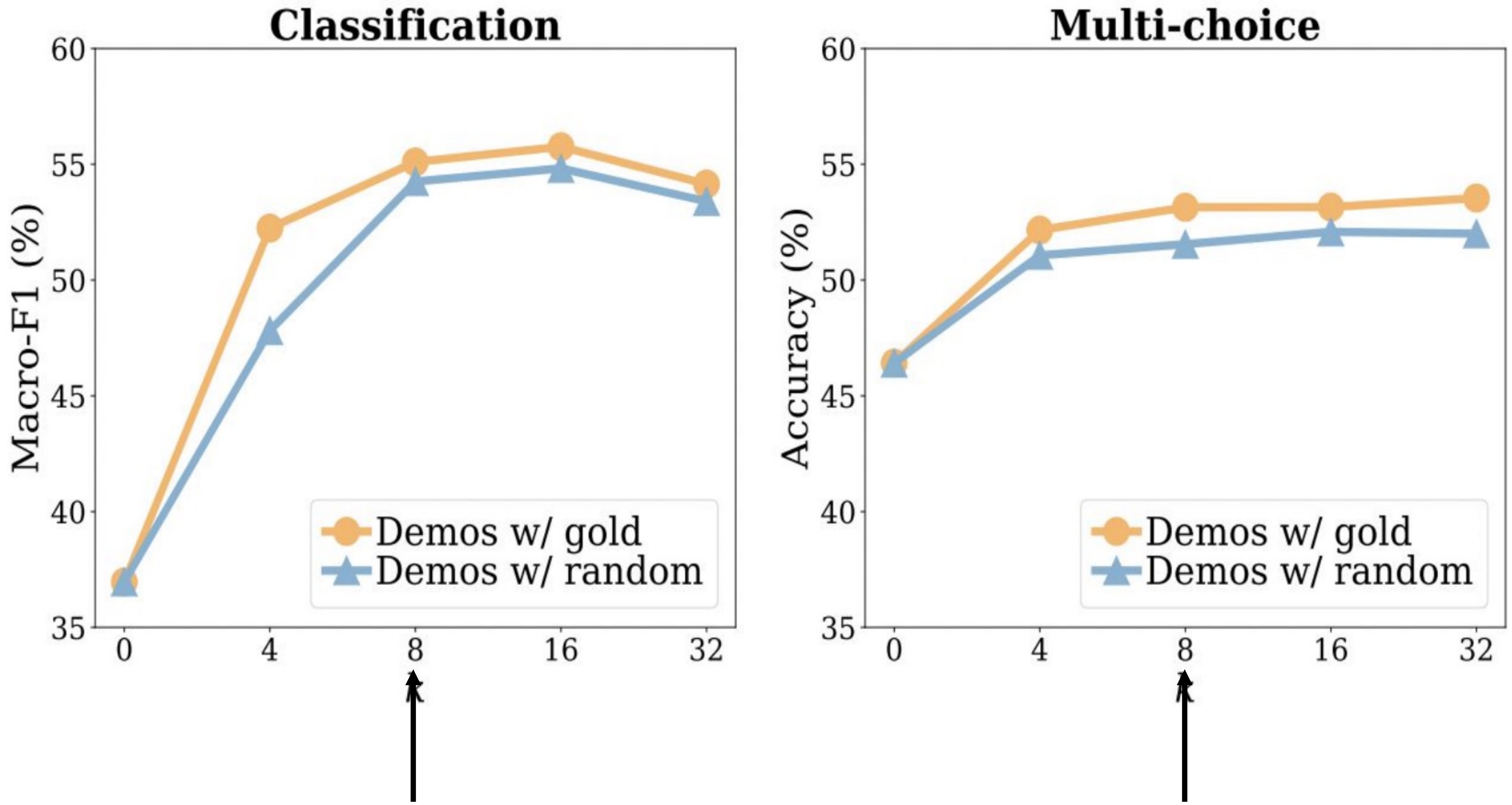


Ablations on varying numbers of examples (k) in the prompt

Performance drop from using gold labels to using random labels is **consistently small** across varying number of examples (k), ranging from 0.8–1.6%

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Results



Ablations on varying numbers of examples (k) in the prompt

More examples even with random labels improves model performance except beyond a threshold

Model performance does not increase much as k increases when k ≥ 8 even for gold labels

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Using Better Templates

Dataset	Type	Example
Tweet_eval-hate	Minimal	The Truth about #Immigration \n {hate|non-hate}
	Manual	Tweet: The Truth about #Immigration \n Sentiment: {against|favor}

Example of minimal and manual templates

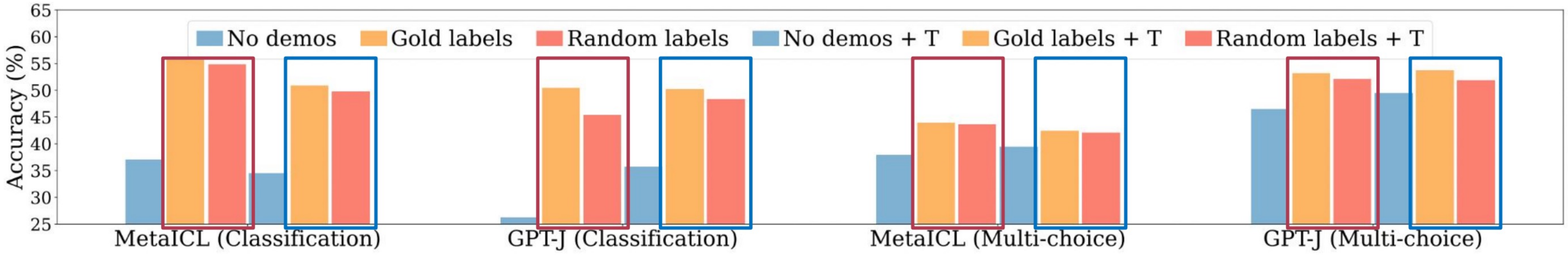
- **Minimal** templates follow a conversion procedure (dataset-agnostic)
- **Manual** templates are written in a dataset-specific manner

Note

Measure whether the results of using random labels is consistent when using manual templates

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Results

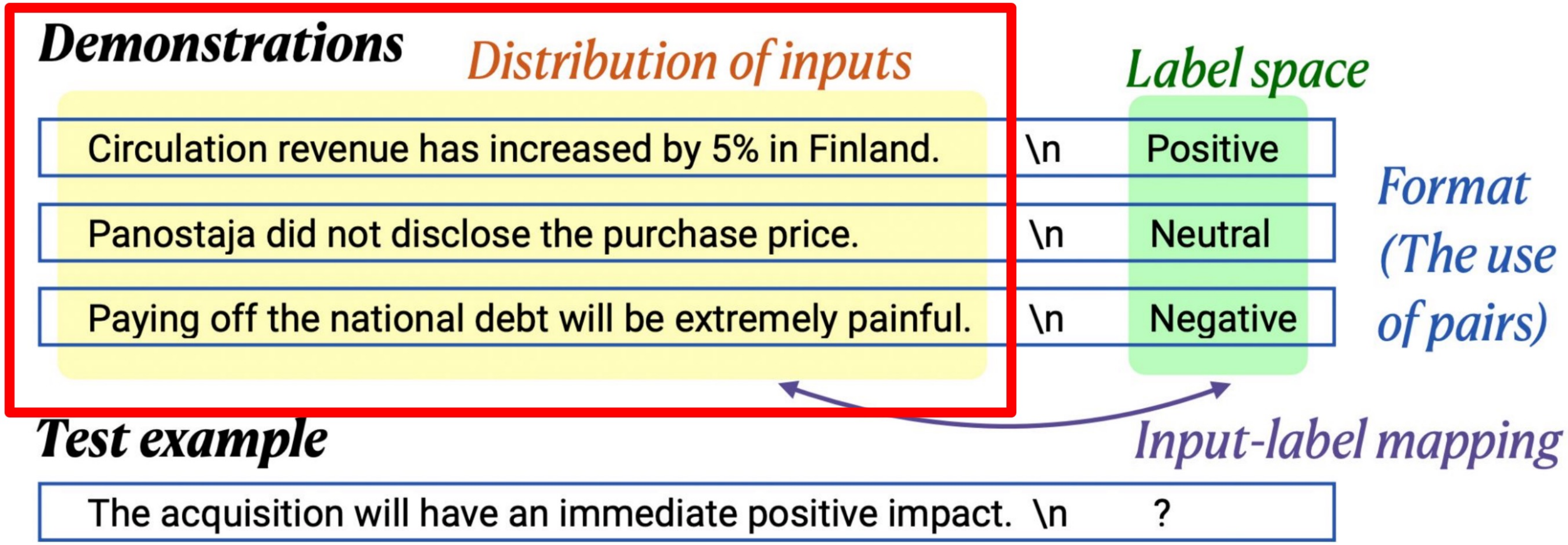


Results with minimal templates and manual templates. '+T' indicates that manual templates are used.

Random labels still minimally hurt performance with manual templates

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Distribution of Inputs



Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Using out-of-distribution input text

Circulation revenue has increased by 5% in Finland.	Positive
Panostaja did not disclose the purchase price.	Neutral
Paying off the national debt will be extremely painful.	Negative
The company anticipated its operating profit to improve.	??

LM

Positive — Correct

Prompt with in-distribution sentences

Colour-printed lithograph. Very good condition.	Neutral
Many accompanying marketing ... meaning.	Negative
In case you are interested in learning more about …	Positive
The company anticipated its operating profit to improve.	??

*Randomly Sampled from CC News

LM

Positive — Correct

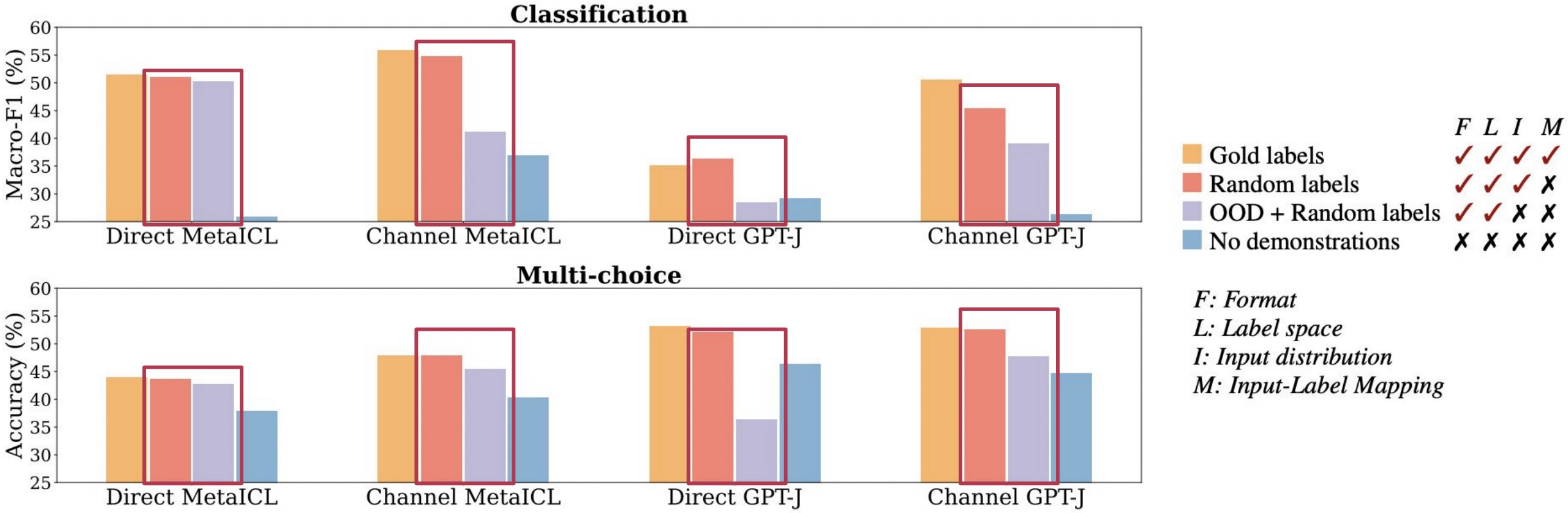
Prompt with out-of-distribution sentences

Note

Input sentences are randomly sampled from an external corpus, replacing the input from the downstream task training data

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Seeing in-distribution inputs improves performance

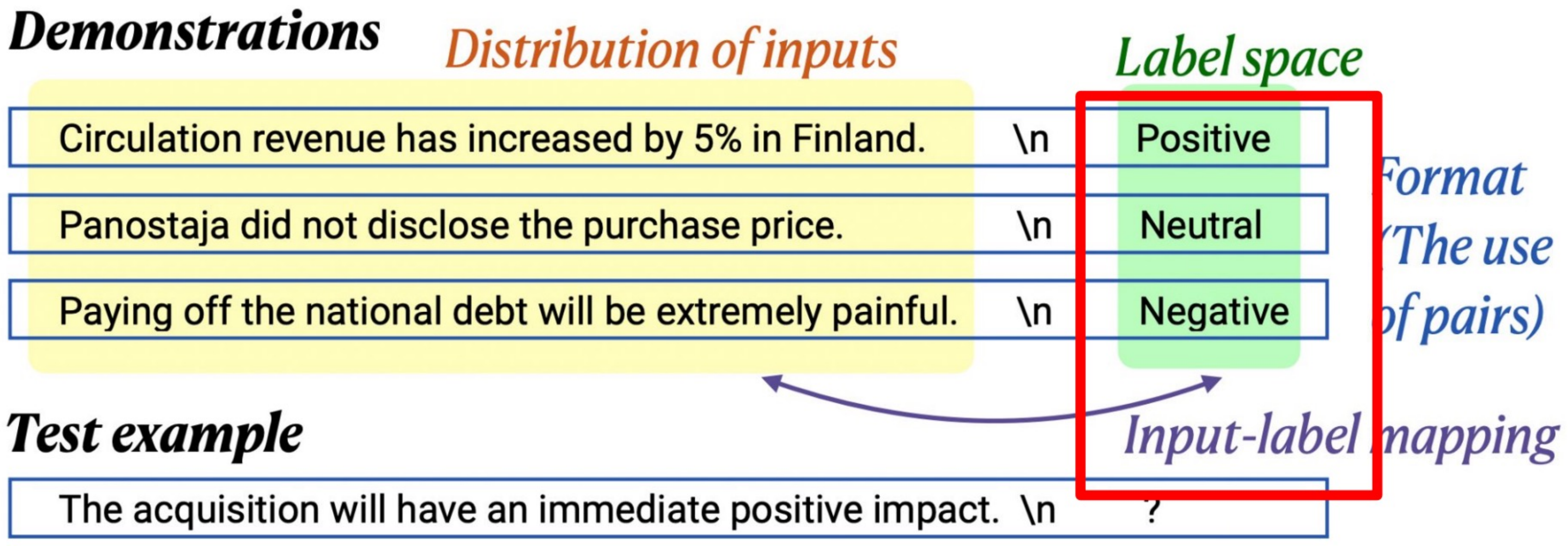


Results of using out-of-distribution input sentences

Random sentences result in performance **decreases of up to 16% absolute** compared to using inputs from training data

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

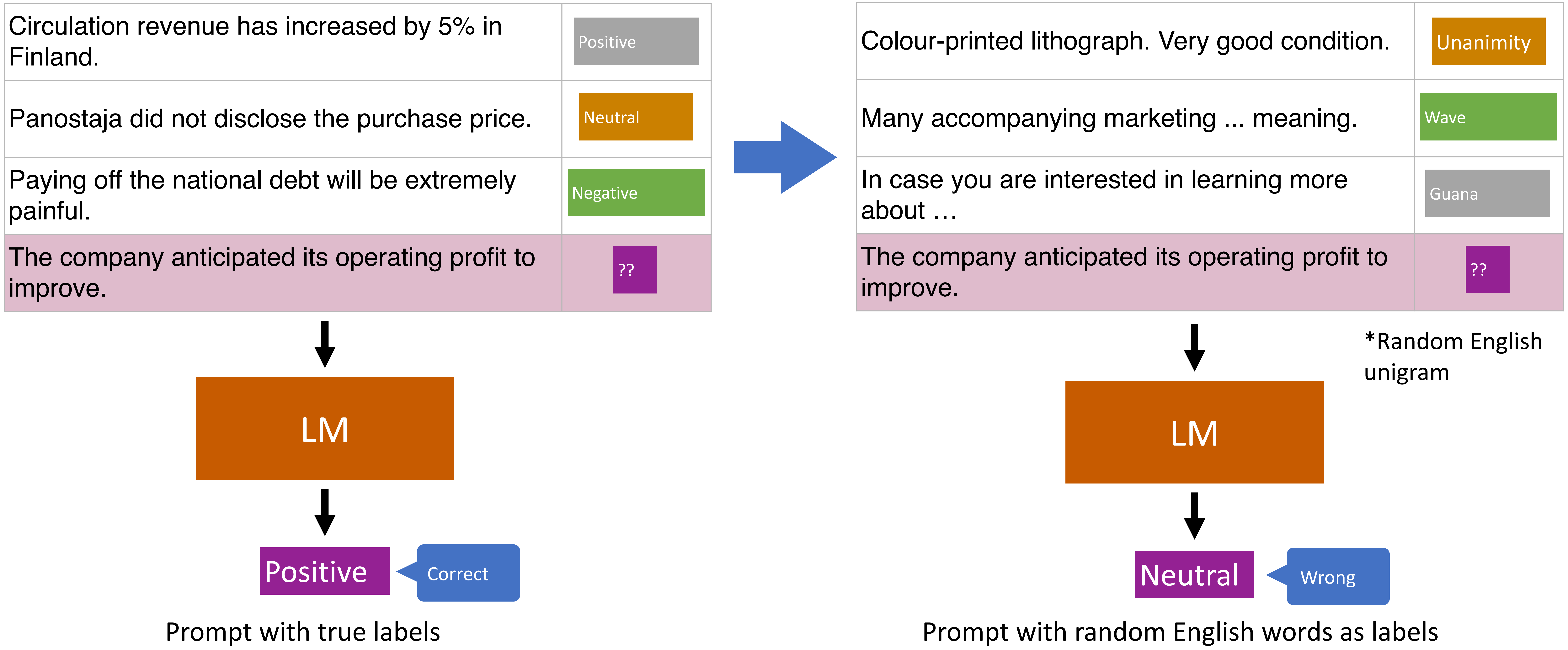
Label Space



Evaluate the importance of the label space

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Using random labels from an incorrect label space

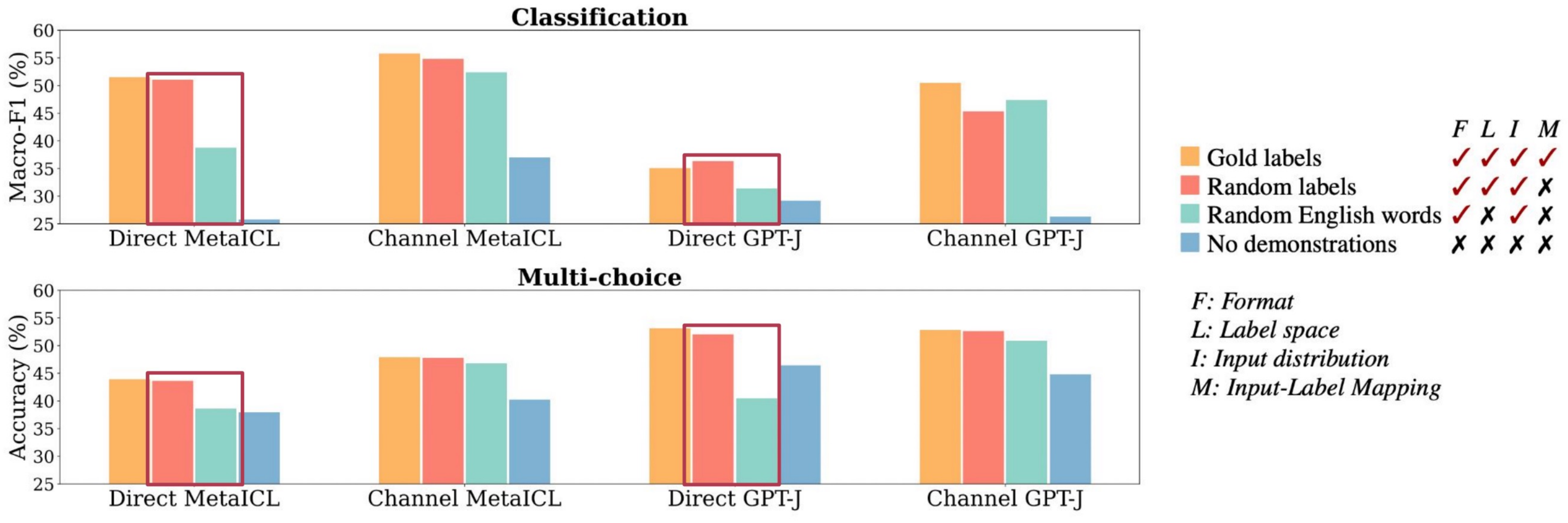


Note

1. Sample a random subset of English words with same size as set of truth labels
2. Labels are replaced with words randomly drawn from this subset

https://ai.stanford.edu/blog/understanding-incontext/

Seeing correct label space is important

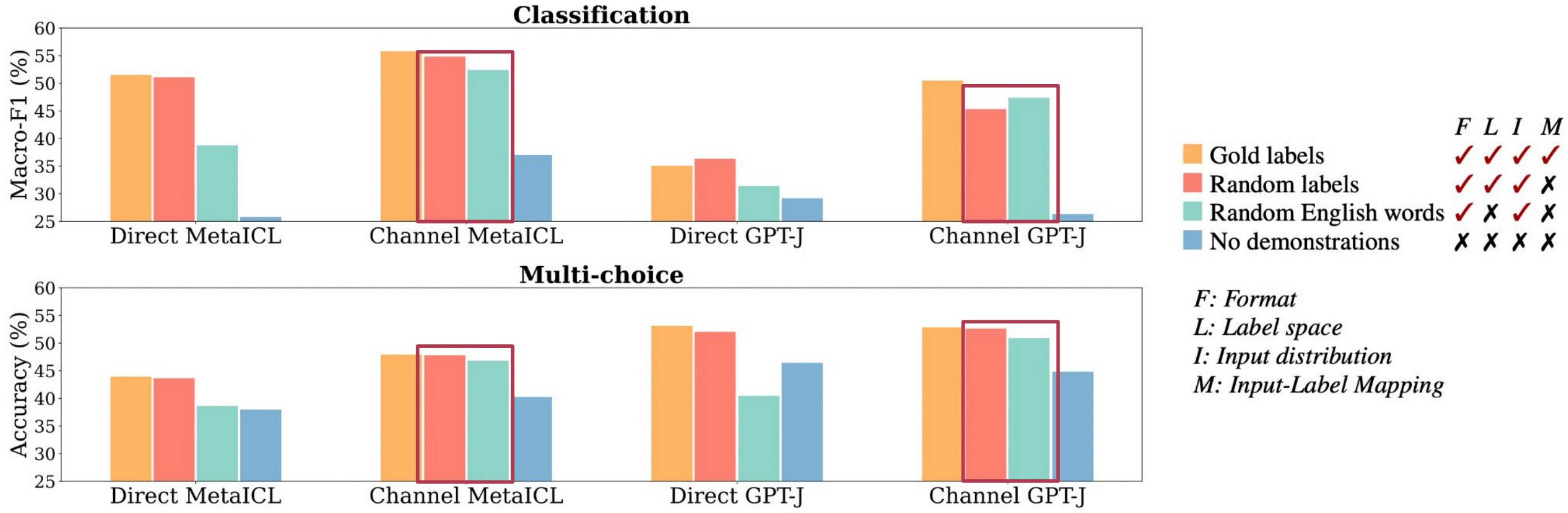


Results of using random English words as labels

Labels not in the correct label space result in **performance decreases of up to 16% absolute** in **direct models**

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Seeing correct label space is important

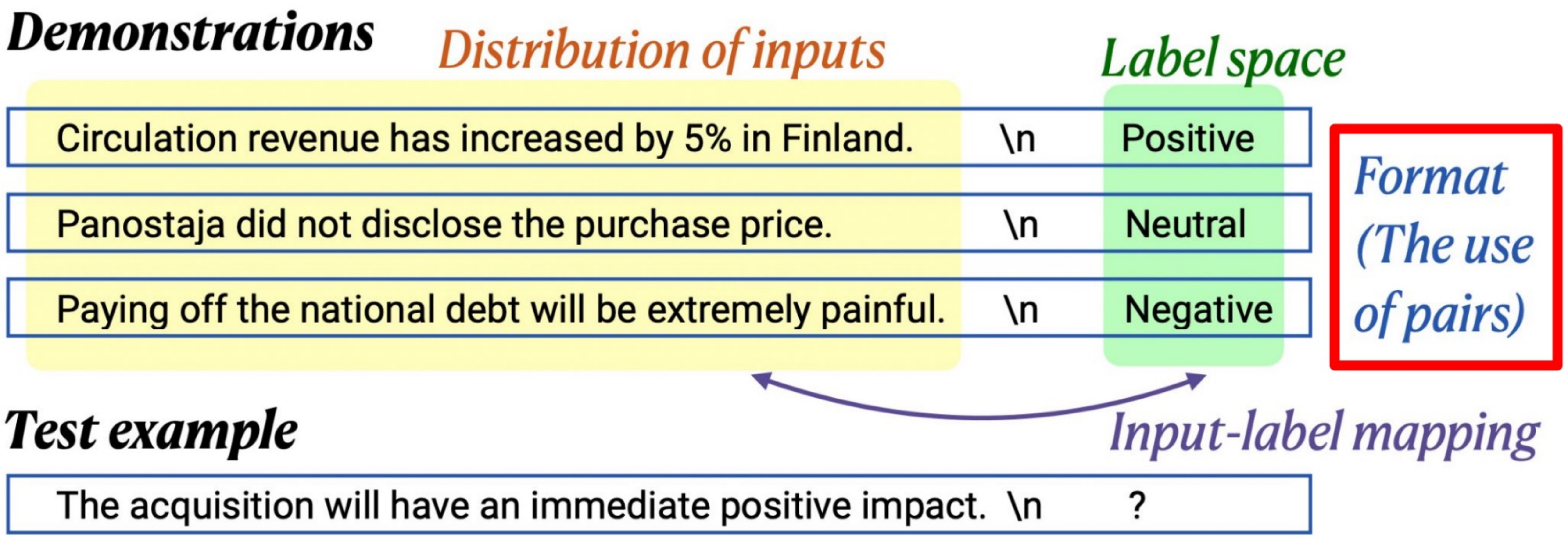


Results of using random English words as labels

Labels not in the correct label space result in **performance decreases of up to 2% absolute** in **channel models**

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Format



Evaluate the importance of pairing an input sentence with a label

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Changing the input-label format

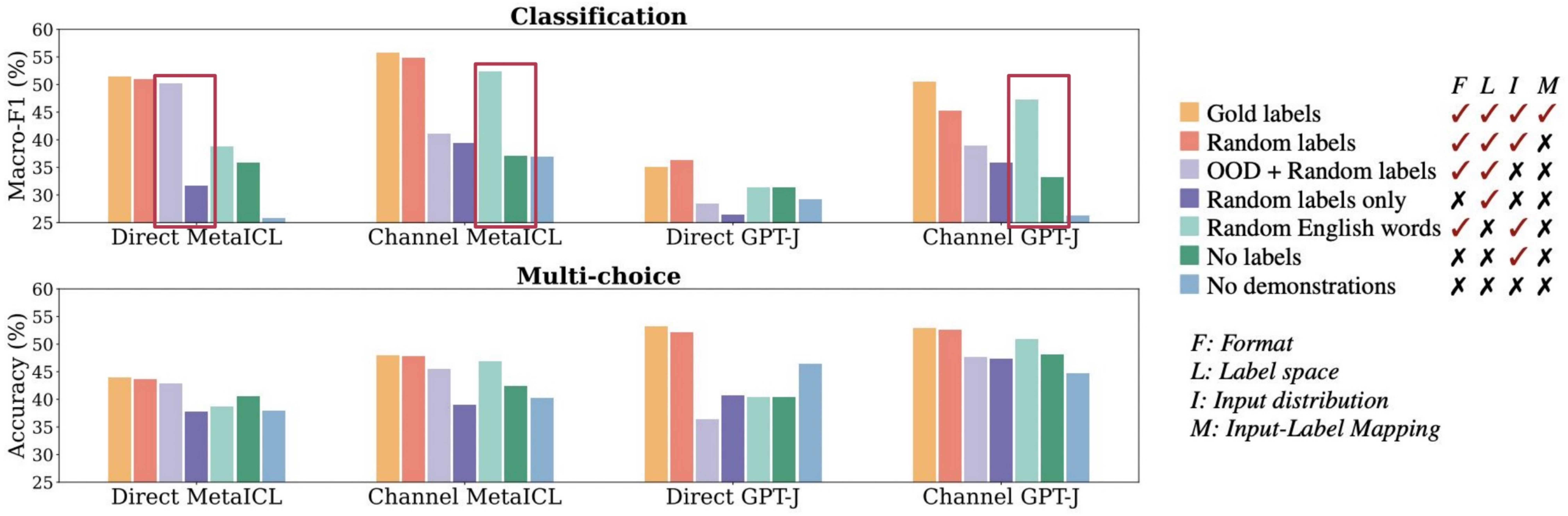
Demos w/o labels	(*Format* ✗ *Input distribution* ✓ *Label space* ✗ *Input-label mapping* ✗) Circulation revenue has increased by 5% in Finland and 4% in Sweden in 2008. Panostaja did not disclose the purchase price.
Demos labels only	(*Format* ✗ *Input distribution* ✗ *Label space* ✓ *Input-label mapping* ✗) positive neutral

Examples with only inputs (top) and only labels (bottom)

Feed in examples with **no labels** and **with labels only**

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Keeping the input-label format for demonstrations is vital for performance



Results of feeding in only inputs and only labels

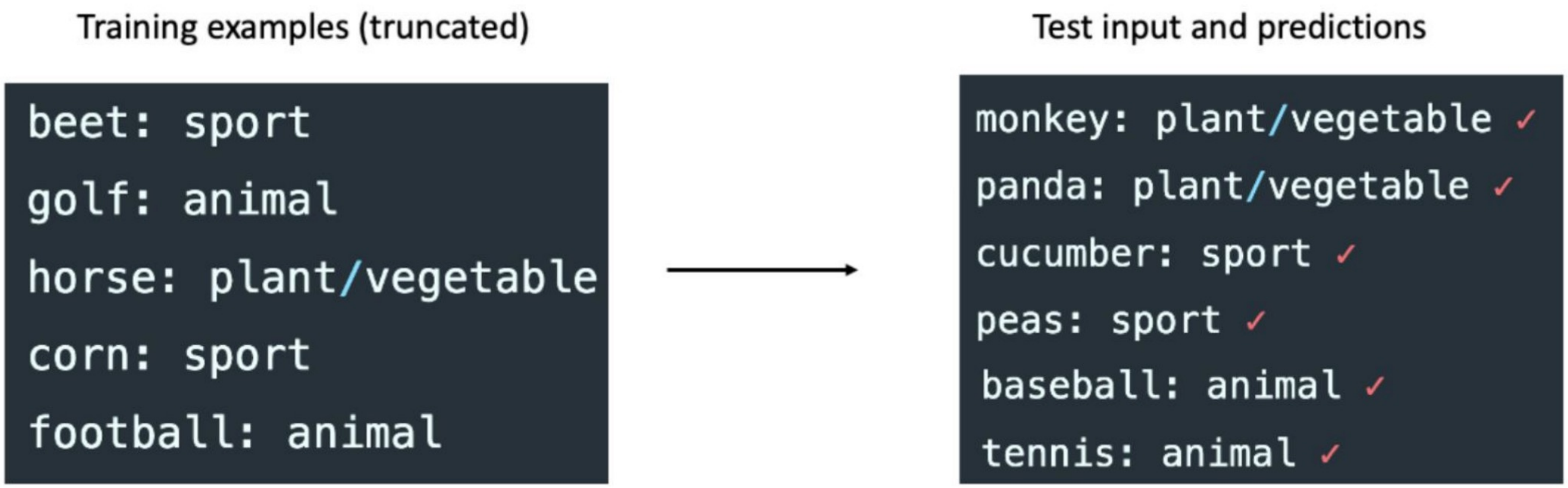
Using **out-of-distribution inputs** and **random English words** as labels is better than only keeping **one part of the format** or having no demonstrations

Min et al., Rethinking the Role of Demonstrations: What Makes In-Context Learning Work, 2022

Rethinking the Role of Demonstrations: Summary

- Having correct input-output pairs do not matter as much as long as we know the correct label space.
- Retaining the format (input-output pairs) whether by using (OOD + random labels) or (in-dist sentences + random English words) also decently improves performance.
- Connection to the Bayesian inference framework
  - all the components of the prompt are providing “evidence” to enable the model to better infer (locate) concepts that are learned during pretraining.

However…



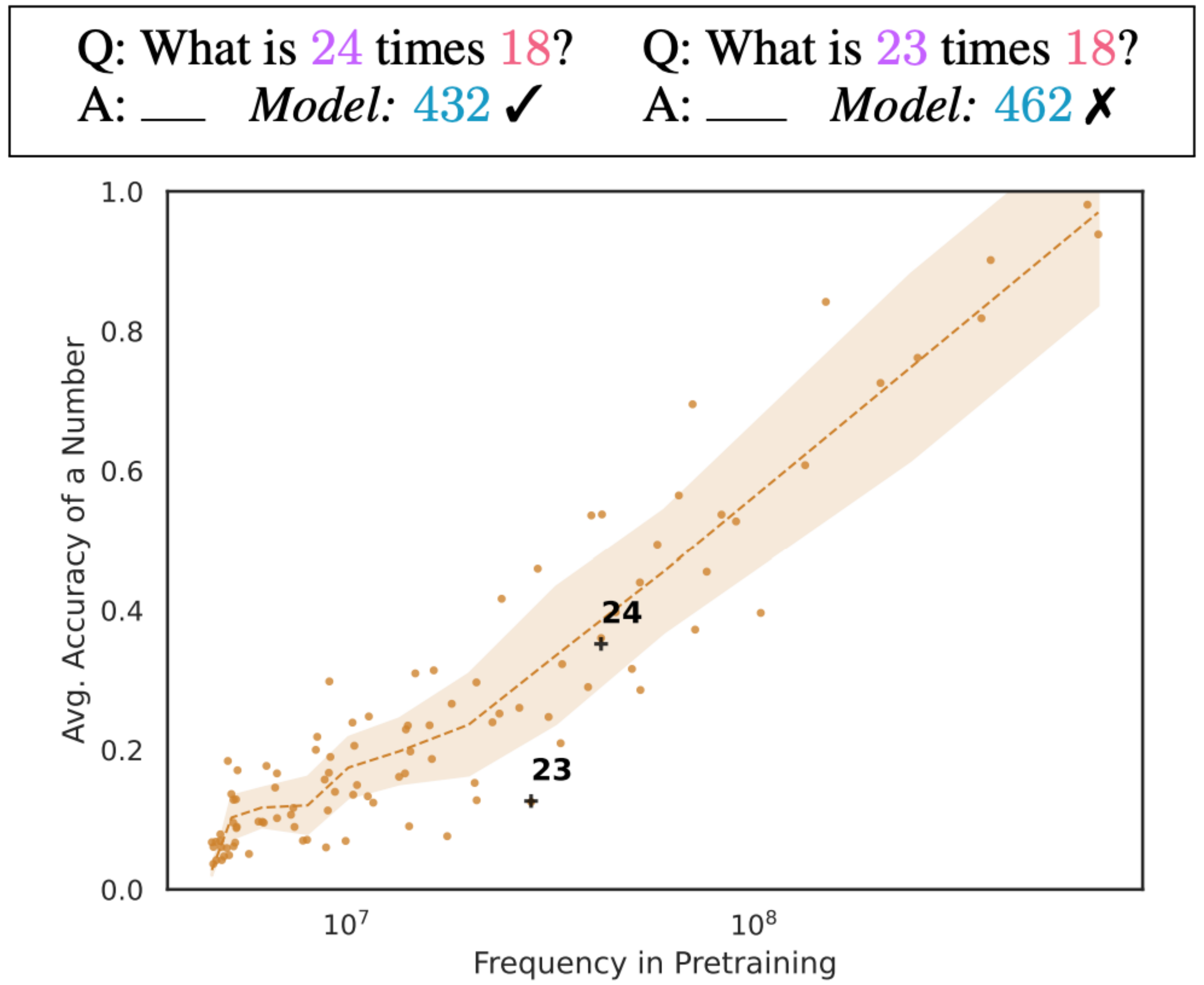
An example synthetic task with unusual semantics that GPT-3 can successfully learn. A modified figure from Rong.

https://ai.stanford.edu/blog/understanding-incontext/

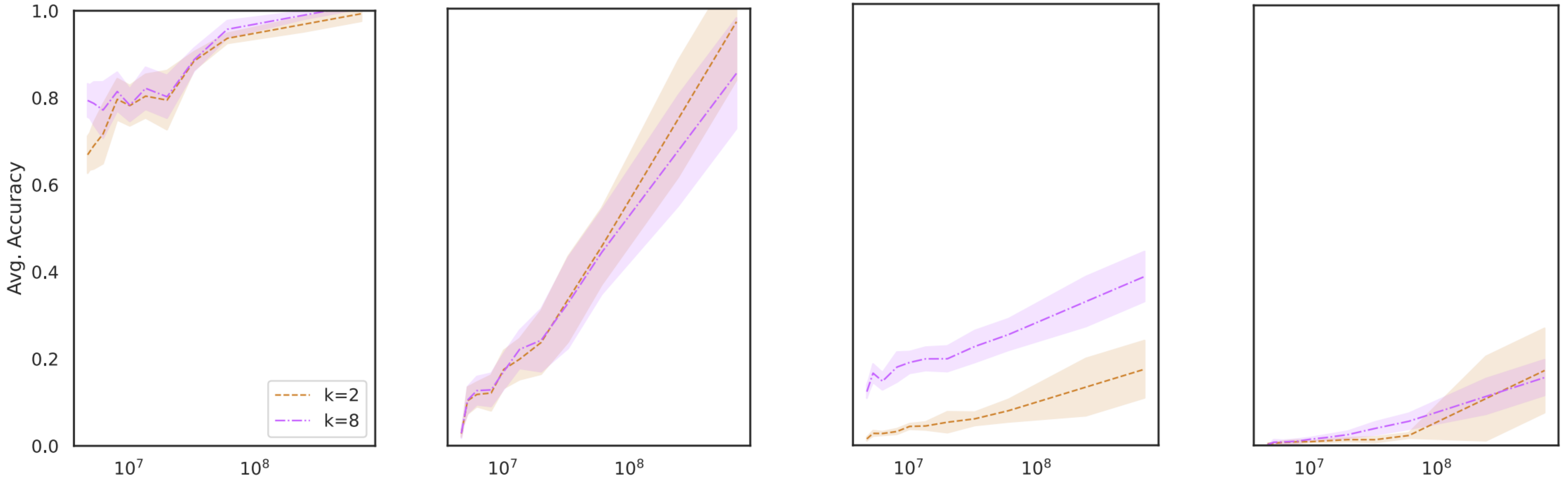
More directions to explore

- **Understanding pre-training data** for in-context learning.
- Understanding model performance on **“unseen” tasks**
- Extending the framework to incorporate **task descriptions** as part of the prompts
- Capturing effects from **model architecture** and training
- **Variable length demonstrations**
  - i.e. k is different in each example

Pretraining data



- In-context learning performance is highly correlated with term frequencies during pretraining



(a) Arithmetic-Addition (b) Arithmetic-Multiplication (c) Op.Inference-Addition (d) Op. Inference-Mult.

Razeghi, Impact of Pretraining Term Frequencies on Few-Shot Numerical Reasoning, 2022

Prompt design

- ICL is inherently unstable
- A surge of methods that search for robust and high-performing prompts:
  - Template search
    - all these methods require a high-quality validation set to do prompt selection or optimization
  - In-context example search

Wu et al., Self-Adaptive In-Context Learning: An Information Compression Perspective for In-Context Example Selection and Ordering, 2022

Choice of demonstrations (examples)

KATE

- ICL is powerful and versatile
- However, its performance depends heavily on the choice of examples.
- How to better select in-context examples?
    - retrieve examples that are semantically-similar to a test sample to formulate its corresponding prompt.

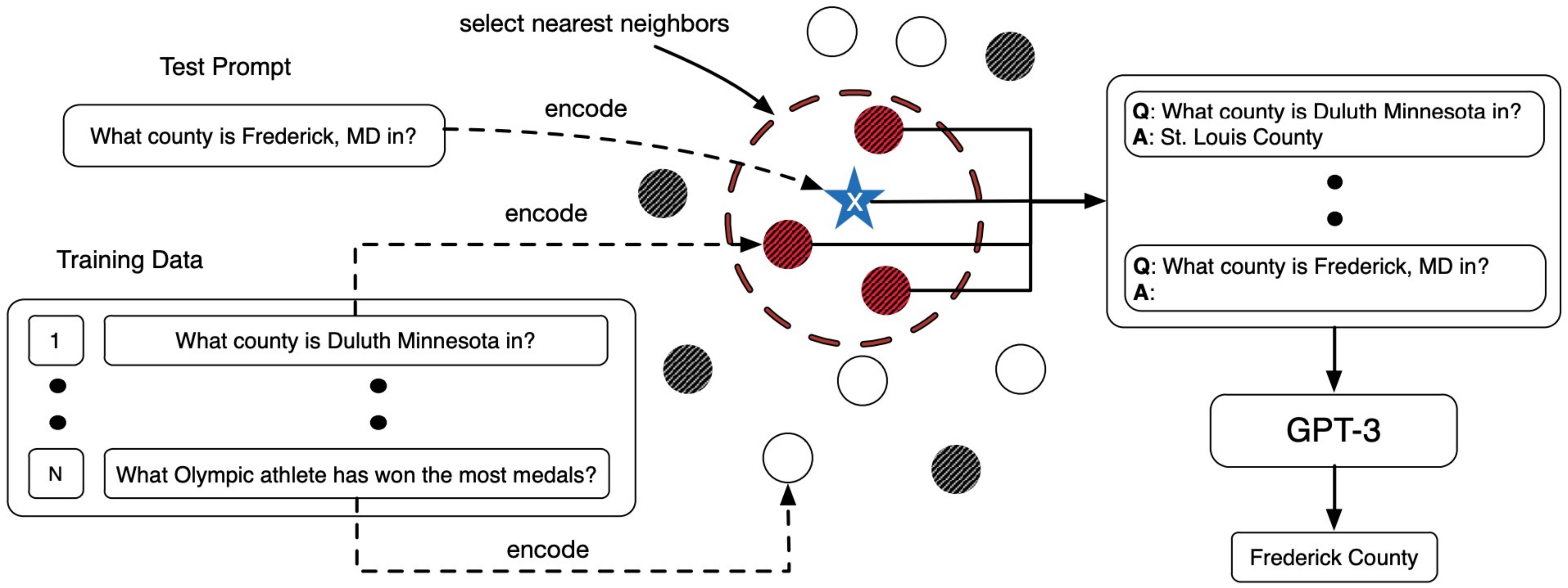
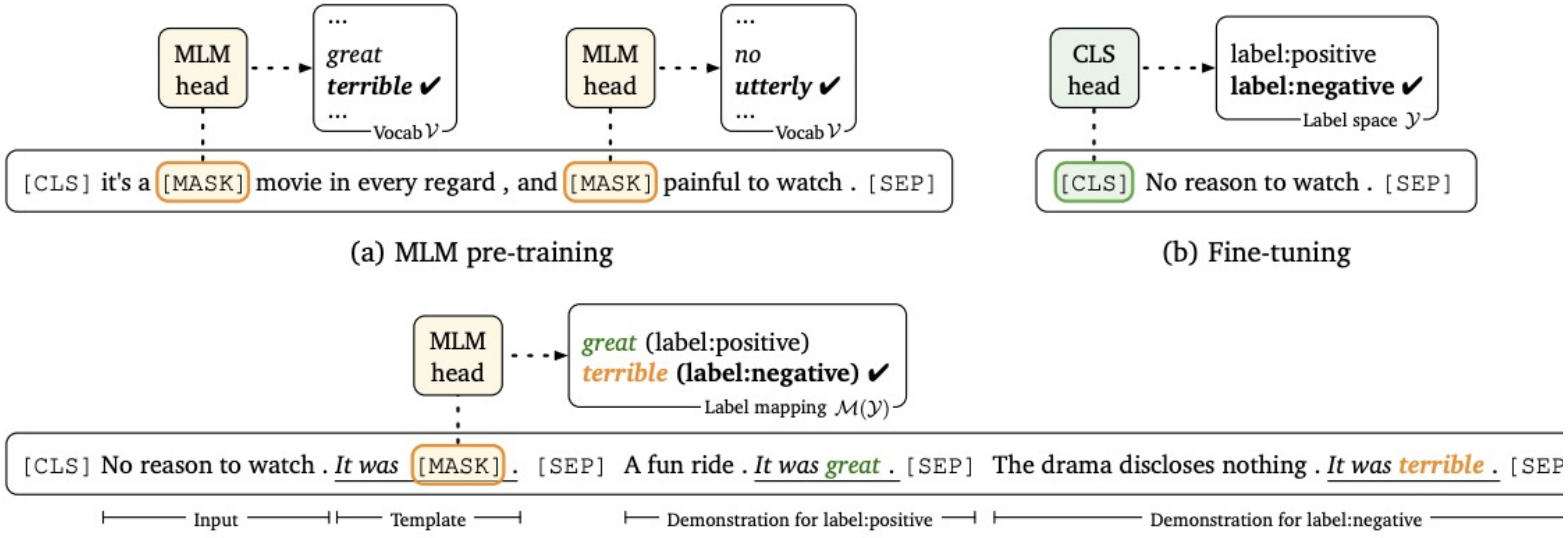


Figure 2: In-context example selection for GPT-3. White dots: unused training samples; grey dots: randomly sampled training samples; red dots: training samples selected by the k-nearest neighbors algorithm in the embedding space of a sentence encoder.

Liu et al., What Makes Good In-Context Examples for GPT-3?, DeeLIO 2022

Recap: Other few-shot learning methods

- Prompt-based finetuning by verbalizers (e.g. LM-BFF)
  - LM-BFF re-uses the pre-trained weights and does not introduce any new parameters. It reduces the gap between pretraining and fine-tuning



(a) MLM pre-training

(b) Fine-tuning

(c) Prompt-based fine-tuning with demonstrations (our approach)

Gao et al., Making Pre-trained Language Models Better Few-shot Learners, ACL 2021

Ways to adapt to new tasks

- Zero-shot learning
  - by task description through Prompt
  - T0: Multi-task training for zero-shot performance
- Few-shot learning
  - In-context learning
  - Verbalizer (i.e. a label word mapping)
- Lightweight Fine-tuning
  - Prompt tuning (Lester et al., 2021)
  - Prefix tuning (Li and Liang, 2021)
  - Adapter (Houlsby et al. 2019) and LoRA (Hu et al., 2021)
- Fine-tuning for human-aligned language models (later in the course)

Questions