

# FINETUNED LANGUAGE MODELS ARE ZERO-SHOT LEARNERS

Benjamín Escobar

# EL PROBLEMA.

Los modelos de lenguaje actuales son muy buenos en "few-shot learning" pero tienen bajo desempeño en "zero-shot learning"

# ¿QUE ES ZERO SHOT LEARNING?

Zero-shot learning se refiere a cuando se testea un modelo con elementos de clases con las que **no se entreno**, tratando de que pueda ejecutar la tarea para la que se diseño de igual manera en estas.

## Ventajas del Zero-Shot Learning:

Se necesitan  
menos datos  
clasificados

Modelos más  
flexibles

Maneja  
casos/categorías  
raras

# STATE-OF-THE-ART.

Hay varias estrategias usadas actualmente para Zero-shot learning:

- Se usan atributos para describir los elementos de cada clase y facilitar las tareas de elementos no vistos.
- Modelos generativos que generen ejemplos de las clases no vistas.
- Modelos de lenguaje pre-entrenados como BERT y las variantes de GPT.

# EL MODELO PROPUESTO

- Se usó un modelo pre-entrenado de 137B de parámetros.
- Se realizó un fine-tuning con más de 60 datasets de NLP divididos en 12 clusters (grupos) con el mismo tipo de tarea

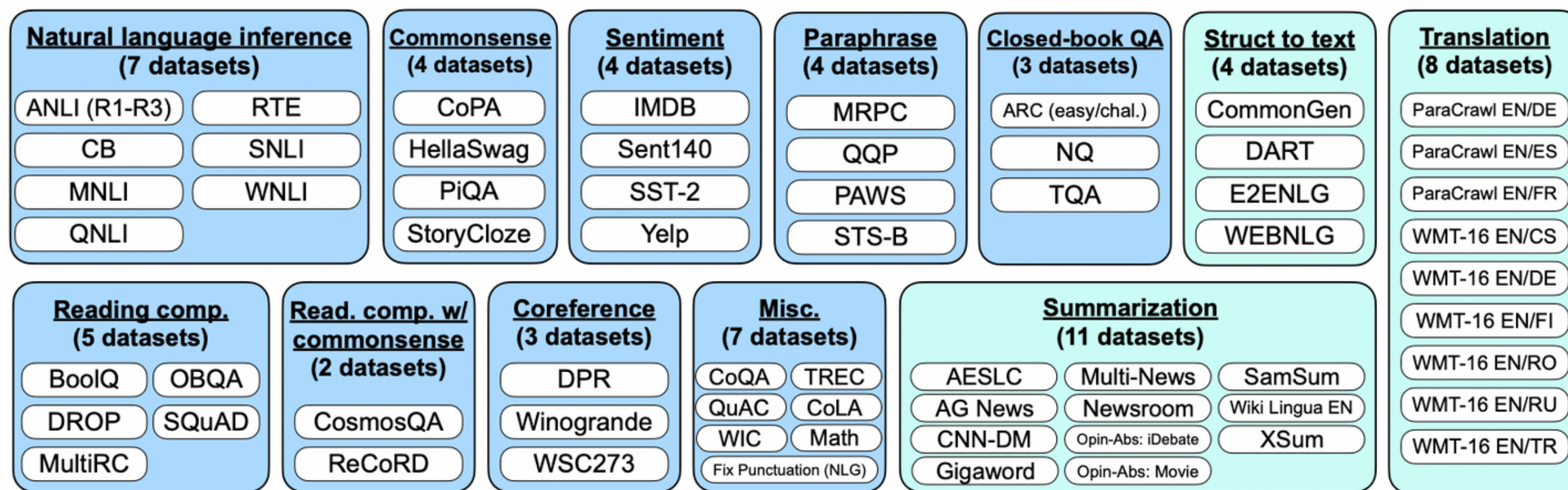
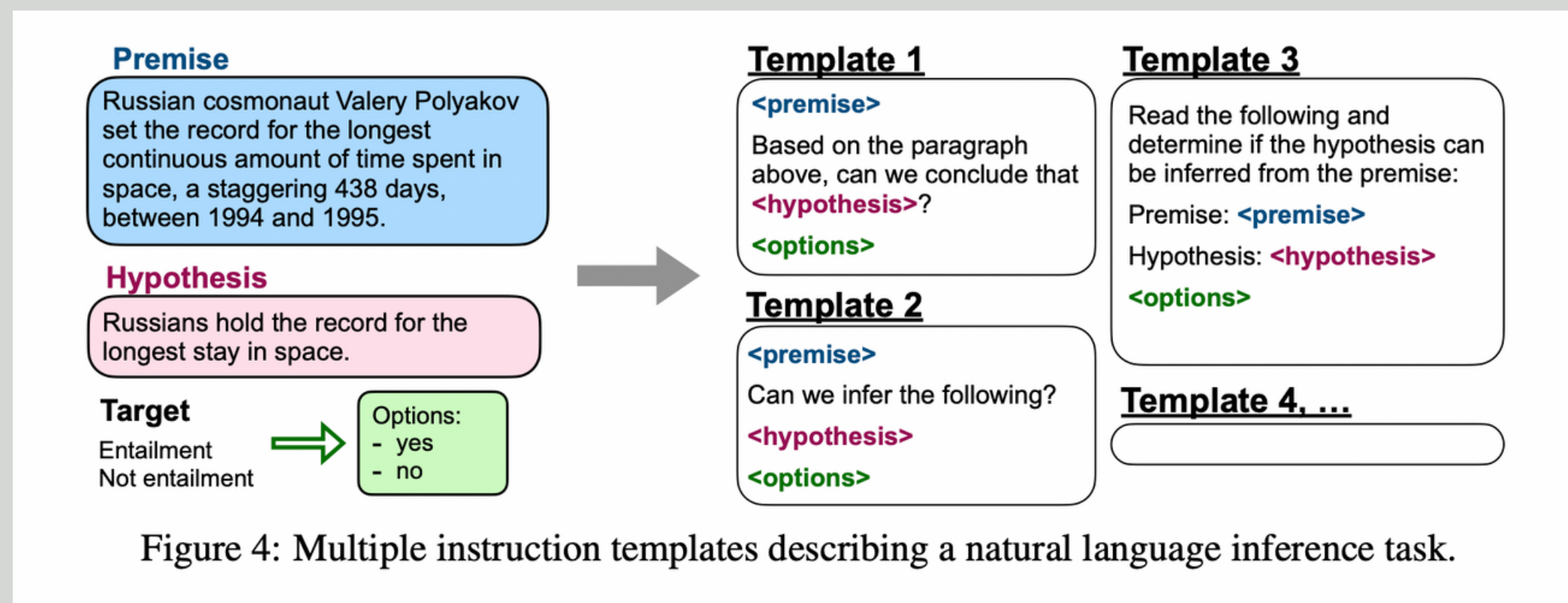


Figure 3: Datasets and task clusters used in this paper (NLU tasks in blue; NLG tasks in teal).

# EL MODELO PROPUESTO

- Se realizaron plantillas manualmente para cada dataset, en el que se verbalizaban las tareas de los datasets de diferentes maneras



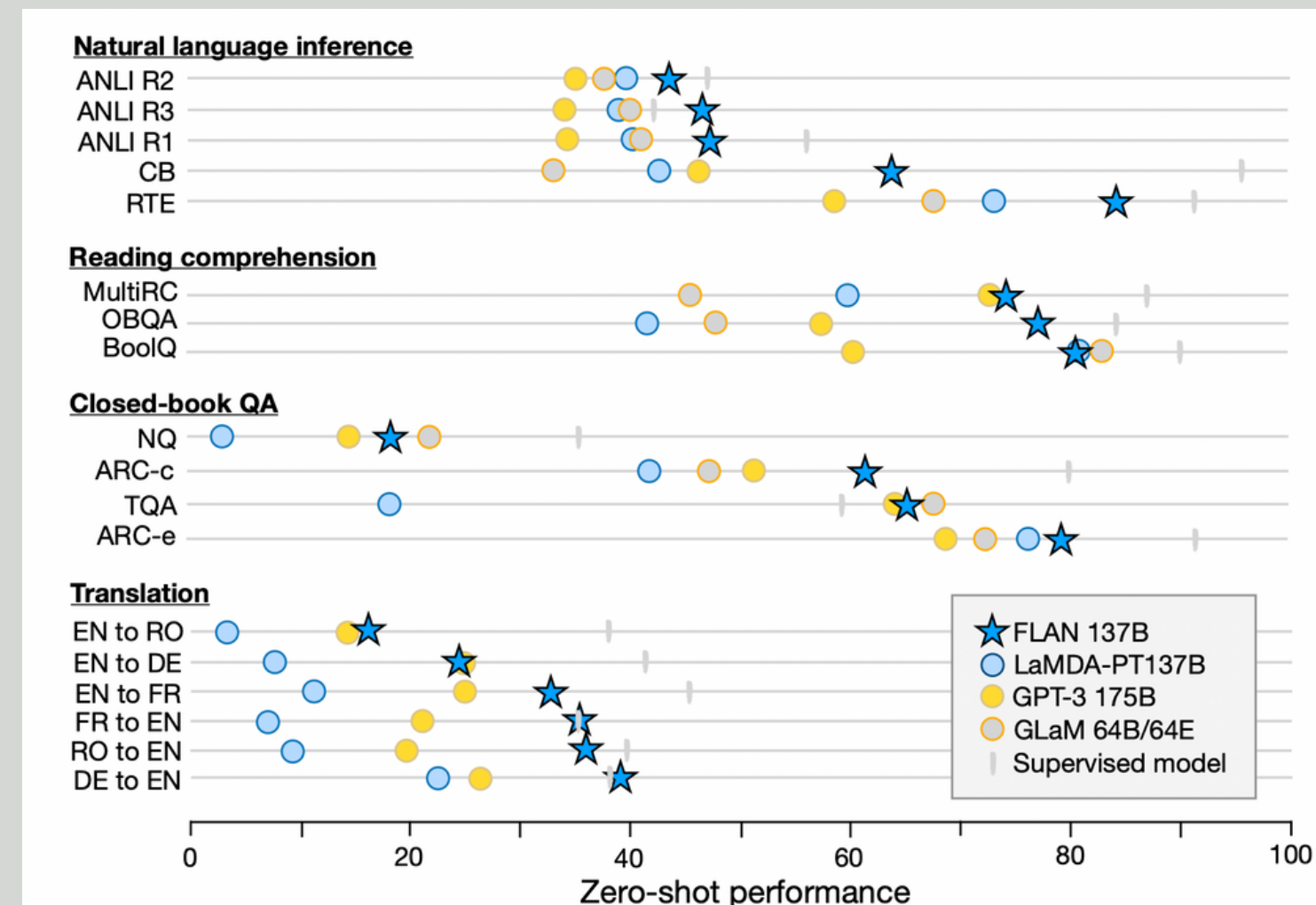


# EL MODELO PROPUESTO

- El modelo se llama FLAN (Finetuned Language Net).
- Para probar su habilidad en clusters no vistos, se hizo el finetunning con distintos clusters excluyendo el que se deseaba testear.
- Luego se evaluó su habilidad de realizar Zero-shot learning con el el cluster no usado.

# EL MODELO PROPUESTO

El modelo fue muy efectivo en varias tareas:



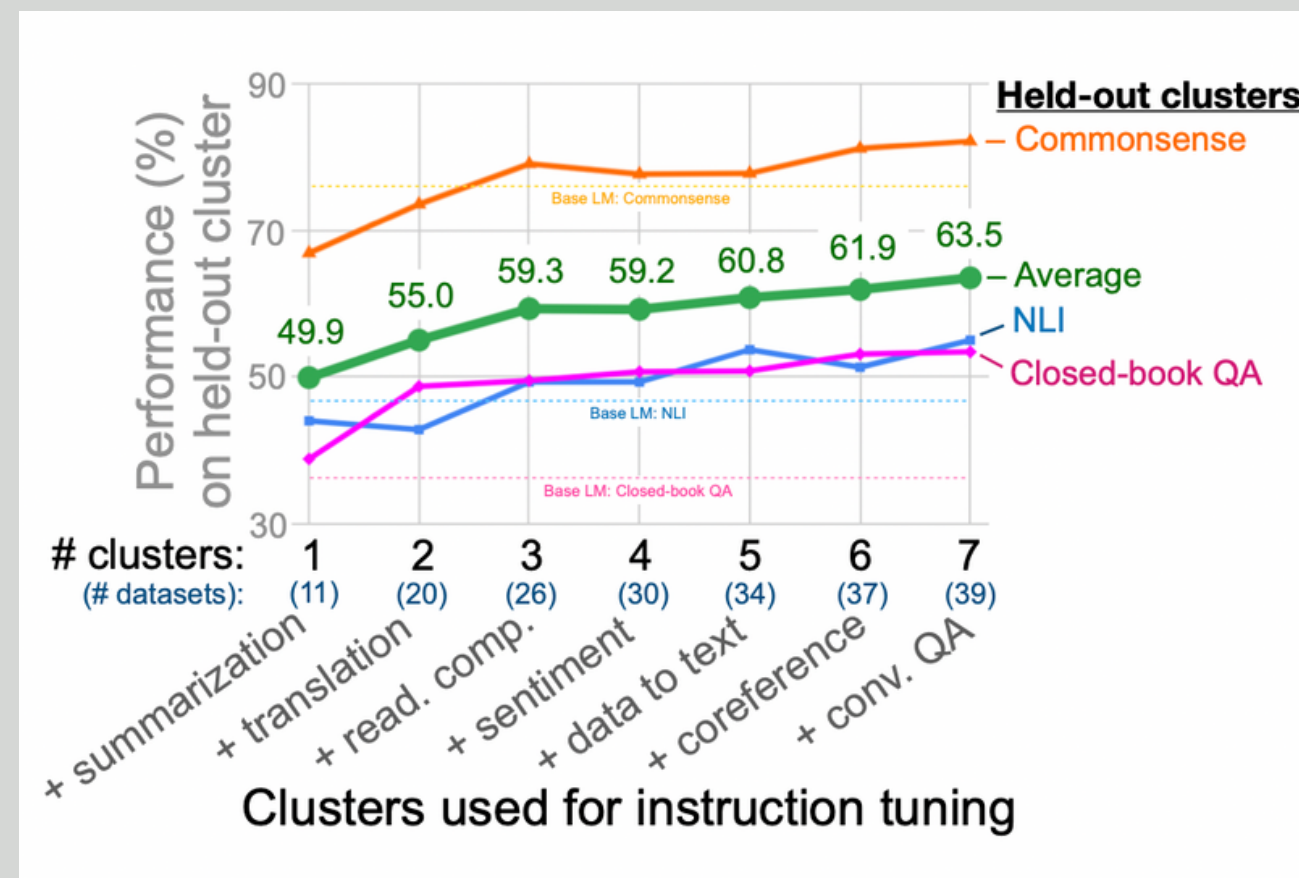


# EL MODELO PROPUESTO

- En las tareas relacionadas con lenguaje tuvo malos resultados.
- Fine-tune no es útil para tareas como completar oraciones, y incluso puede empeorar el modelo pre-entrenado. Hipótesis del paper: Al re-entrenar un modelo que ya podía completar oraciones con más oraciones, los datos pueden ser redundantes y hacer que empeore el desempeño en esos casos.

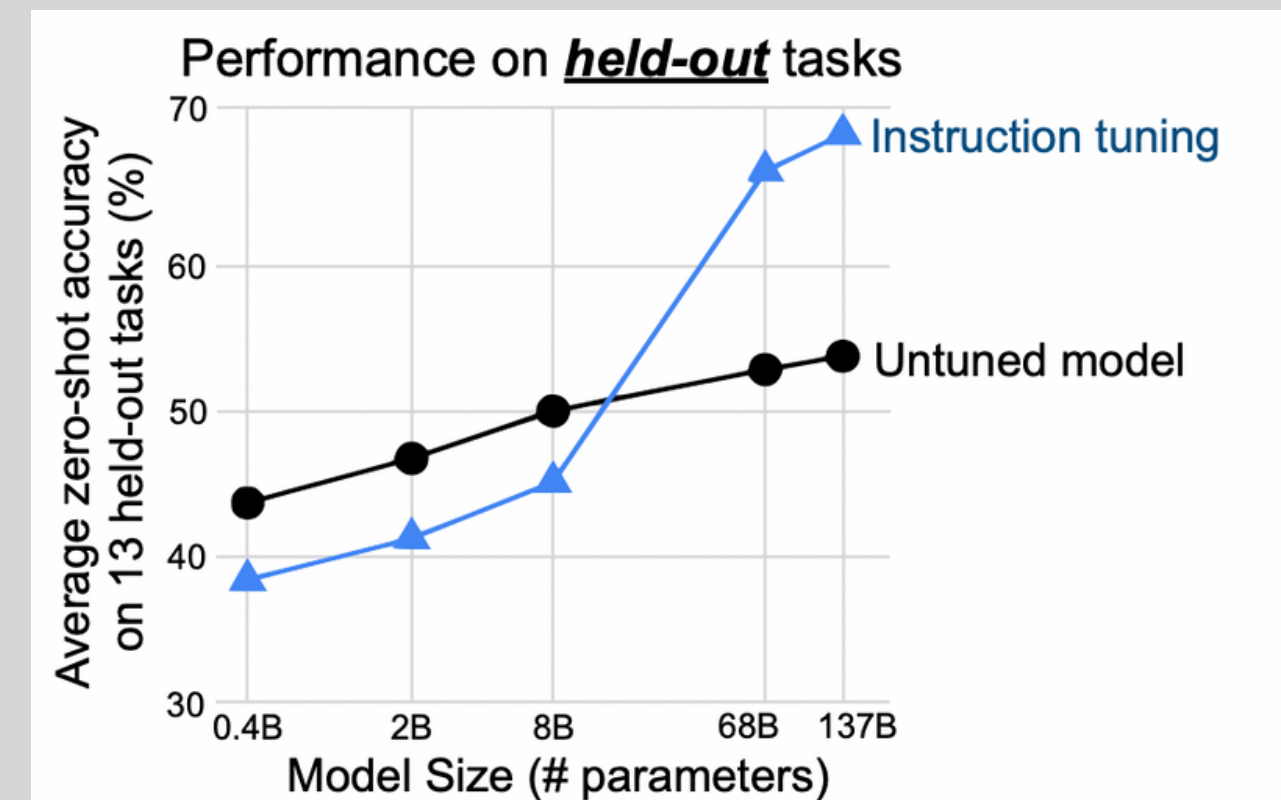
# EL MODELO PROPUESTO

- Entrenar con más clusters mejora el rendimiento.
- Algunos clusters en particular no tienen un beneficio considerable, particularmente el de sentimientos.



# EL MODELO PROPUESTO

- Se probó el método con modelos de distinta cantidad de parámetros.
- Fine-tuning funciona mejor con modelos más grandes.



# CONCLUSIONES

- El modelo usado resulto ser muy efectivo en Zero-Shot Learning comparado con las otras alternativas en varias tareas.
- Tiene defectos en tareas relacionadas con lenguaje como continuar frases.
- En Few-Shot learning tiene un peor desempeño que su competencia en general.

# CRITICA AL PAPER



Esta escrito de una manera que es fácil de entender.



Reconoce las deficiencias y es una buena contribución en mi opinión.



Se requería harto conocimiento externo al Paper.