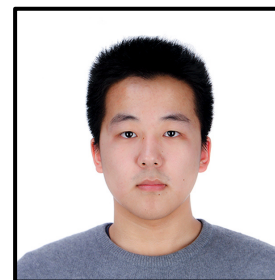
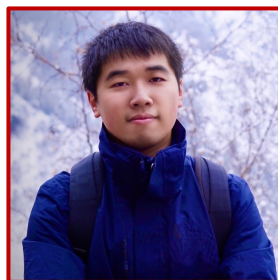


Unsupervised Cross-Task Generalization via Retrieval Augmentation



Bill Yuchen Lin[†] Kangmin Tan[†] Chris Miller[†] Beiwen Tian[‡] Xiang Ren[†]

[†] University of Southern California [‡] Tsinghua University

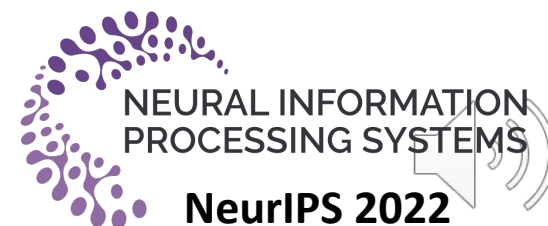
{yuchen.lin, kangmint, millercs, xiangren}@usc.edu



USC

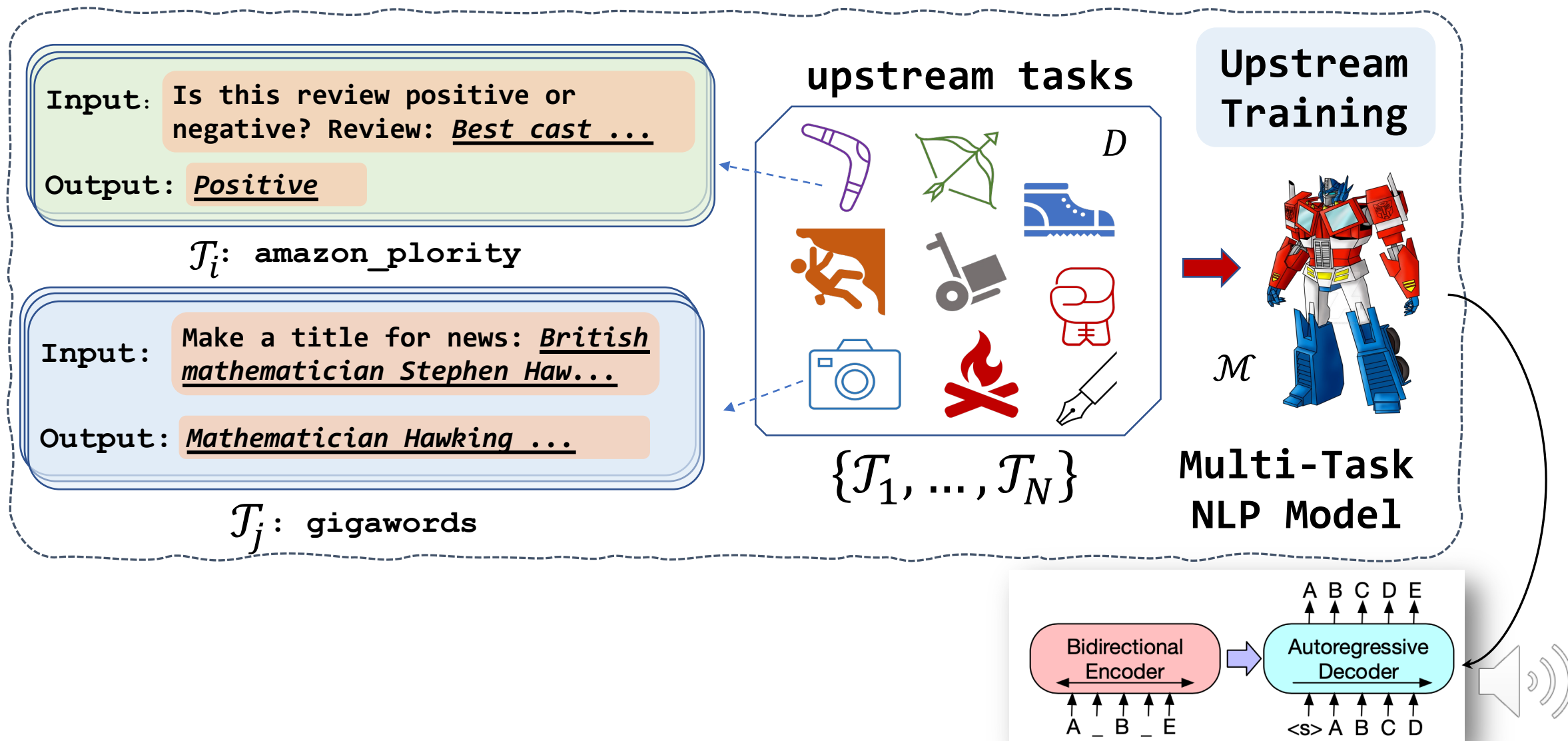


清华大学
Tsinghua University



<https://inklab.usc.edu/ReCross/>

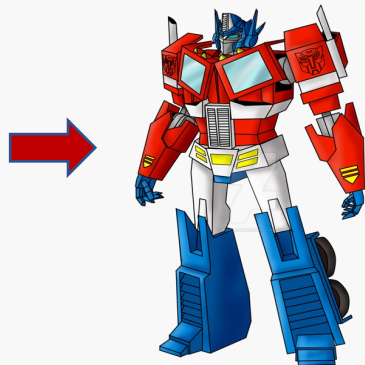
Massively Multi-Task NLP Models



upstream tasks



Multi-Task NLP Model



Sentiment Analysis

Review: We came here on a Saturday night and luckily it wasn't as packed as I thought it would be [...] On a scale of 1 to 5, I would give this a

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?

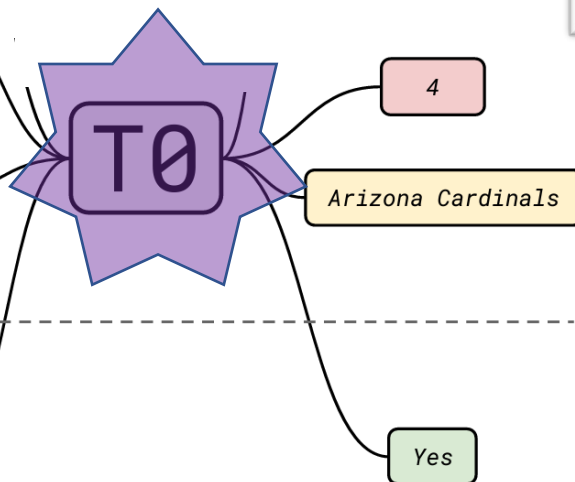
Multi-task training

Zero-shot generalization

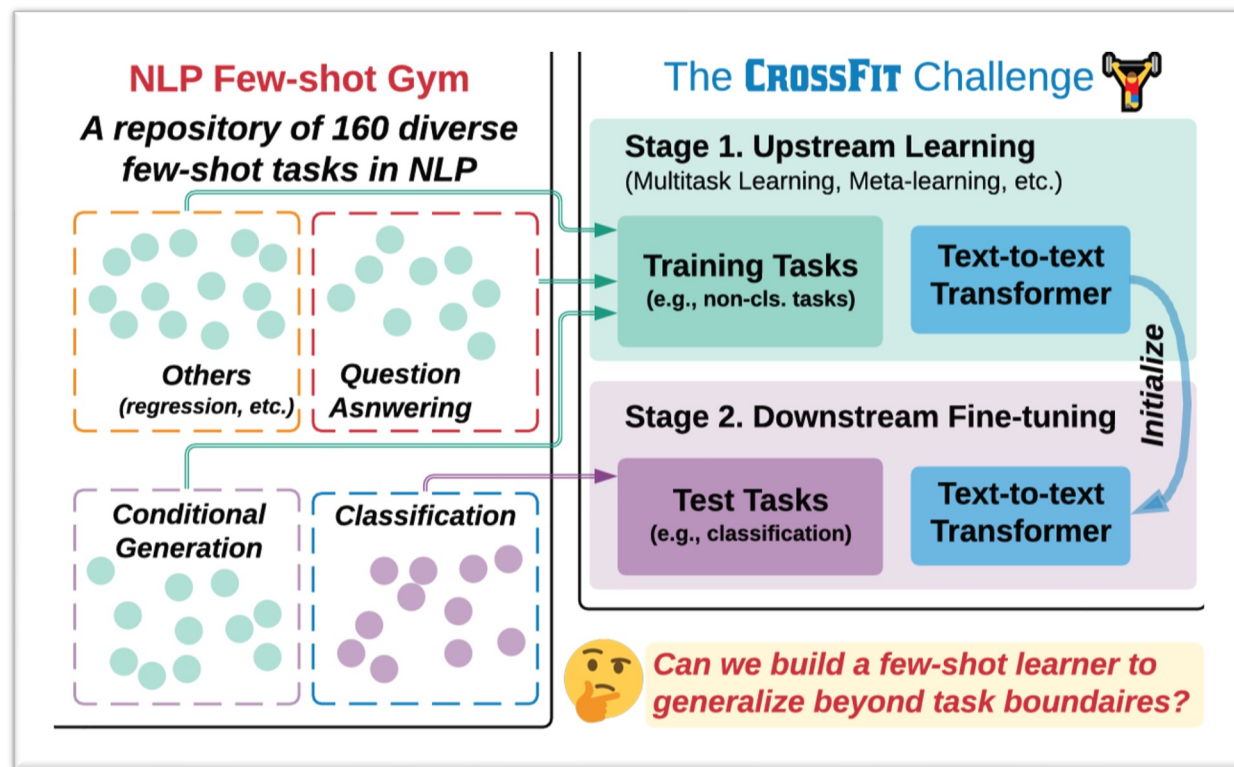
Natural Language Inference

Suppose "The banker contacted the professors and the athlete". Can we infer that "The banker contacted the professors"?

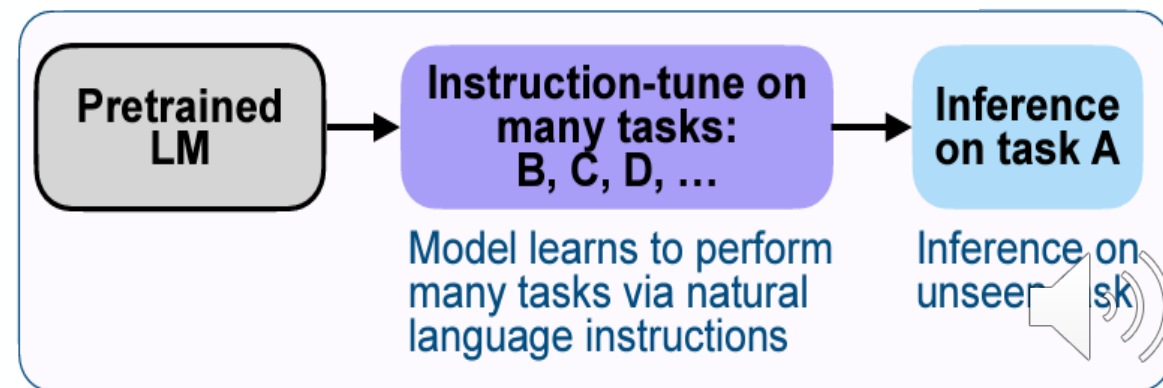
T0 (Sanh et al. 2022)



CrossFit (Ye et al. 2021)



FLAN (Wei et al. 2022)

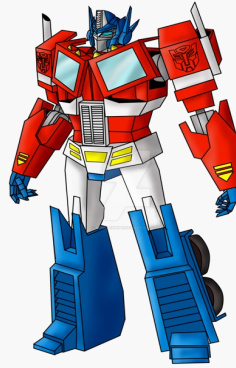


Unsupervised Cross-Task Generalization

upstream tasks



Multi-Task
NLP Model



I want to ski!



Input: *Fred is ... He* In the sentence, does the pronoun "he" refer to Fred? Yes, no?

Input: Does "her" refer to Alice or Mary in this sentence: Alice ... Mary... and her ...

Q_i

u_i : an unseen task w/ a few unlabeled data

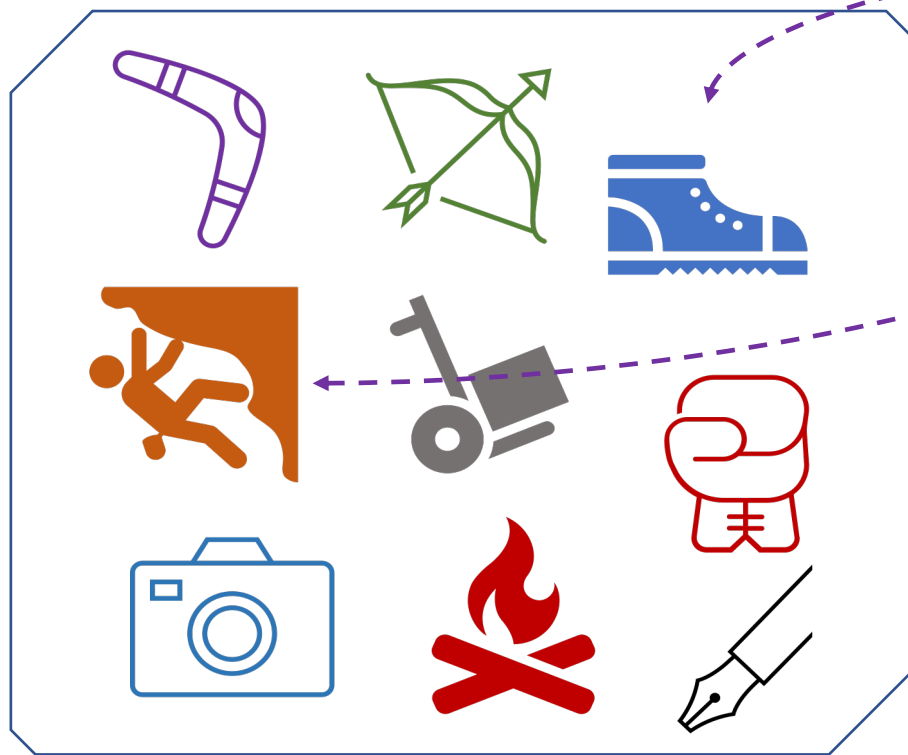


Unsupervised
Cross-Task Generalization



Retrieval Augmentation

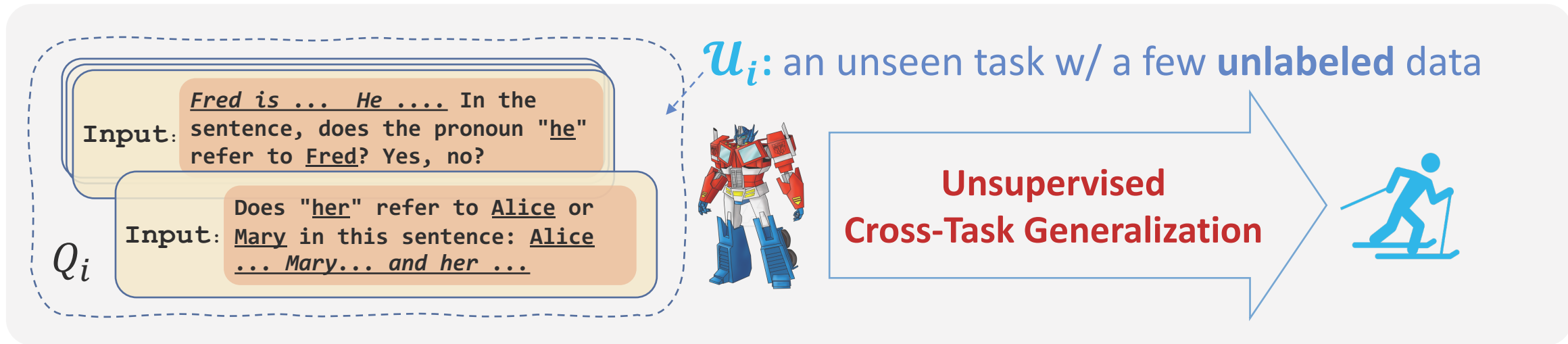
upstream tasks



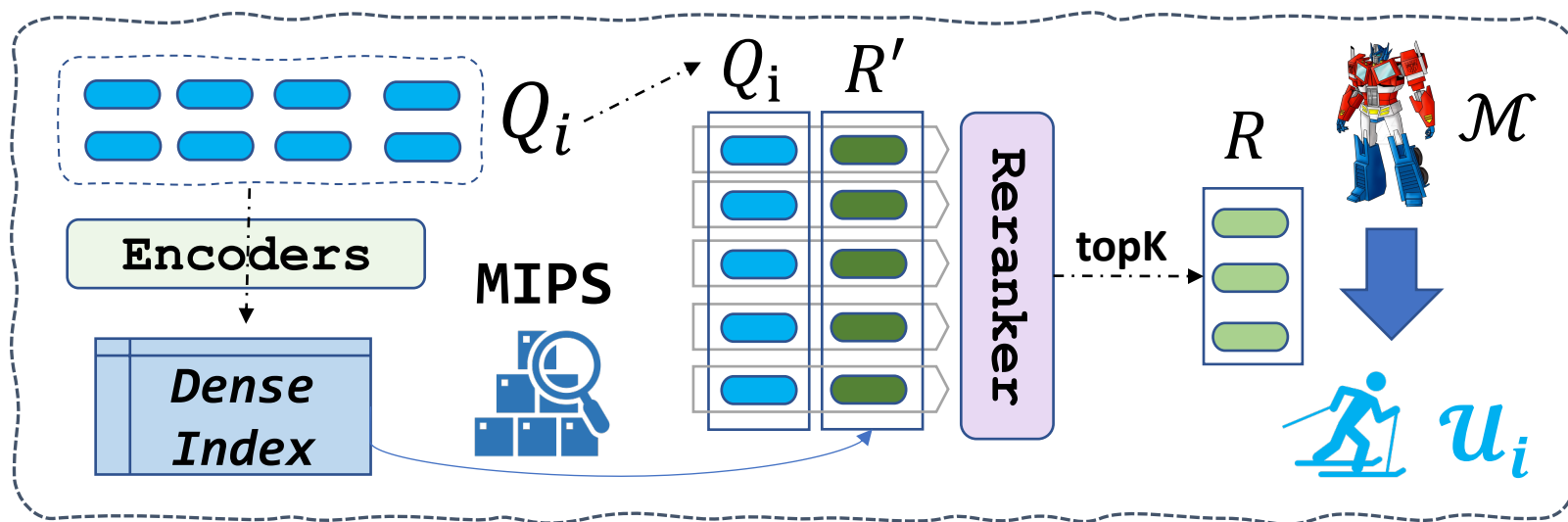
$\mathcal{U}_i$: an **unseen** task w/
a few unlabeled data

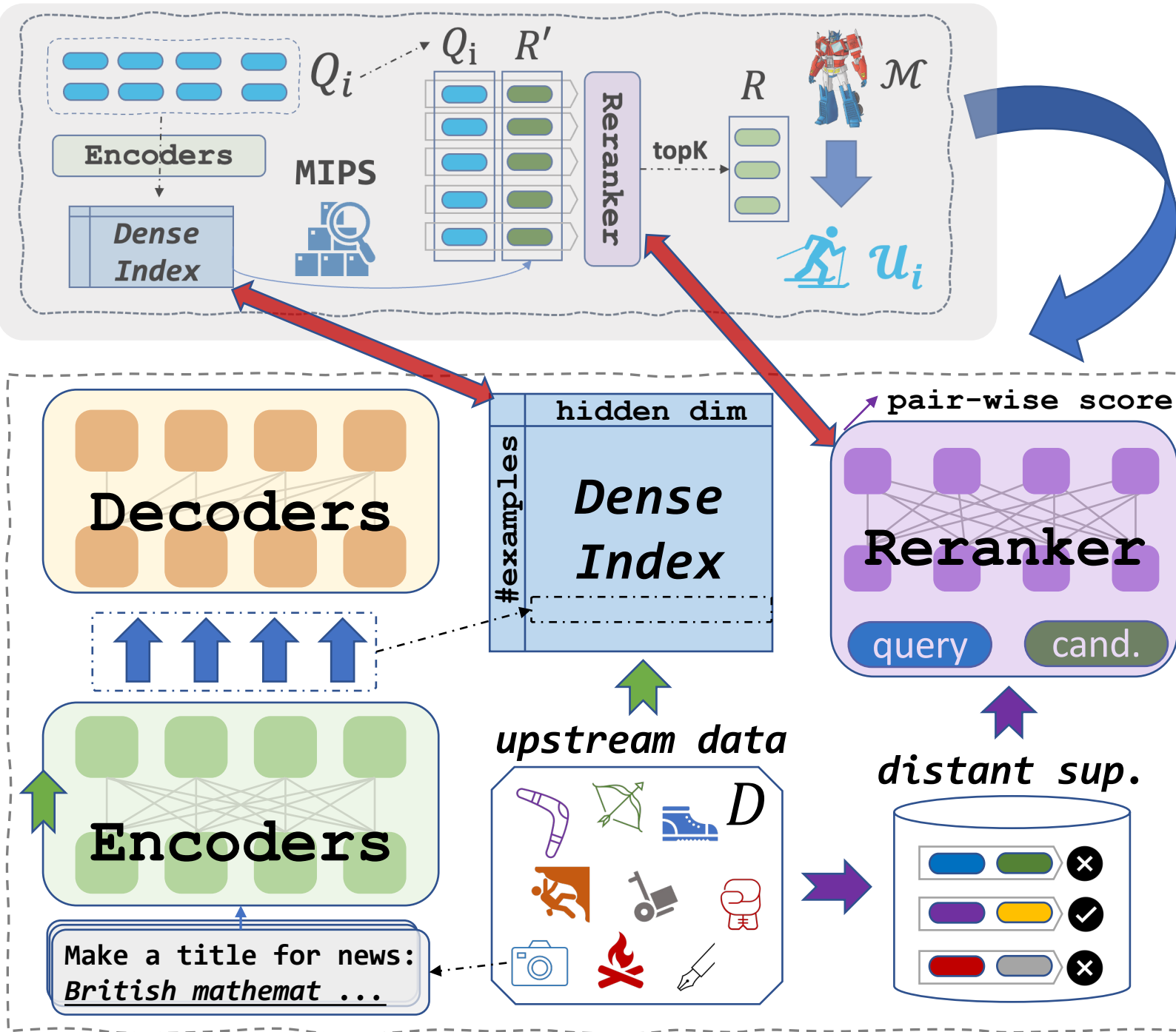
Retrieve some additional data
for “**re-learning**” relevant skills.





The ReCross Framework





Algorithm 1: Distant Supervision Creation

Input: $\mathcal{M}$; D ; $\mathcal{T}_q$

Output: $Z = (Z_q, Z_p, Z_n)$

```

 $D_{\mathcal{T}_q} \leftarrow \{x \in D \mid x \text{ is an example of } \mathcal{T}_q\}$ 
 $Z_q \leftarrow \text{Sample}(D_{\mathcal{T}_q}); H_q \leftarrow \text{Sample}(D_{\mathcal{T}_q})$ 
 $R_Z \leftarrow \text{DenseRetrieve}(Z_q, D)$ 
/* Delete retrieved examples from the same task as queries. */
 $R_Z \leftarrow R_Z.\text{discard}(D_{\mathcal{T}_q})$ 
foreach round do
     $R_Z.\text{shuffle}()$ 
    /* Split retrieved examples into  $n$  groups */
     $\{G_1, \dots, G_n\} \leftarrow R_Z.\text{split}()$ 
    foreach  $G_i$  in  $\{G_1, \dots, G_n\}$  do
         $\mathcal{M}' \leftarrow \mathcal{M}.\text{copy}()$ 
         $\mathcal{M}'.\text{fine\_tune}(G_i)$ 
         $\ell \leftarrow \mathcal{M}'.\text{calc\_loss}(H_q)$ 
        foreach  $x \in G_i$  do
             $\text{scores}[x].\text{append}(\ell)$  /* Score each
                                     example in the group w/ the loss. */
    /* Use mean group score as score for single examples */
    foreach  $x \in R_Z$  do
         $\text{score}[x] \leftarrow \text{mean}(\text{scores}[x])$ 
    /* Sort  $R_Z$  by score in increasing order. */
     $R_Z.\text{sort}(\text{key: score, order: increasing})$ 
     $Z_p \leftarrow \text{First } W \text{ items of } R_Z$ 
     $Z_n \leftarrow \text{Last } W \text{ items of } R_Z$ 

```



Evaluation

1/8 size

Target Task	T0-3B	BART0	Random	SBERT	ReCross [†]	ReCross	Δ
anli_r3	26.00	30.50	35.34 $\pm$ 1.52	32.64 $\pm$ 2.53	36.70 $\pm$ 0.53	35.76 $\pm$ 0.90	5.26
h-swag	34.40	39.40	33.84 $\pm$ 5.59	30.92 $\pm$ 7.82	44.36 $\pm$ 3.07	47.28 $\pm$ 2.95	7.88
cb	53.93	39.64	47.07 $\pm$ 1.25	48.00 $\pm$ 3.28	44.50 $\pm$ 4.20	44.79 $\pm$ 3.36	5.15
wic	45.70	46.70	41.04 $\pm$ 2.18	46.78 $\pm$ 2.22	49.90 $\pm$ 0.50	50.58 $\pm$ 0.24	3.88
wsc	50.00	57.88	52.50 $\pm$ 2.29	52.69 $\pm$ 6.13	59.27 $\pm$ 1.96	61.46 $\pm$ 1.47	3.58
winogrande	47.60	51.10	52.68 $\pm$ 0.83	52.18 $\pm$ 3.20	54.60 $\pm$ 1.35	55.46 $\pm$ 0.88	4.36
arc-chan.	41.30	35.70	33.28 $\pm$ 1.50	37.90 $\pm$ 1.22	37.78 $\pm$ 0.73	38.44 $\pm$ 0.99	2.74
obqa	38.50	34.40	28.72 $\pm$ 2.46	33.28 $\pm$ 1.24	36.98 $\pm$ 1.55	39.58 $\pm$ 2.80	5.18
piqa	45.30	36.10	37.00 $\pm$ 2.71	38.54 $\pm$ 2.17	41.34 $\pm$ 1.75	41.42 $\pm$ 1.02	5.32
squadv2	30.60	32.40	29.86 $\pm$ 5.46	29.46 $\pm$ 0.84	30.26 $\pm$ 1.54	30.58 $\pm$ 1.61	-1.82
All@mean	41.33	40.38	39.13 $\pm$ 2.06	40.24 $\pm$ 1.61	43.57 $\pm$ 0.68	44.53 $\pm$ 0.42	4.15
@median	41.33	40.38	39.93	40.91	43.43	44.31	3.93
@min	41.33	40.38	35.66	38.28	42.65	44.16	3.77
@max	41.33	40.38	40.59	41.76	44.51	45.07	4.69

5 different query sets: Q



Conclusion

- **ReCross**: a retrieval augmentation method for unsupervised cross-task generalization.
 - Two ranking stages: simple & intuitive
 - Significant improvement.
- **Future directions**
 - More rigorous task/example representation for indexing.
 - Parameter-efficient tuning for faster adaptation.
 - Methods to use unlabeled data for guiding re-learning.
- <https://inklab.usc.edu/ReCross/>

