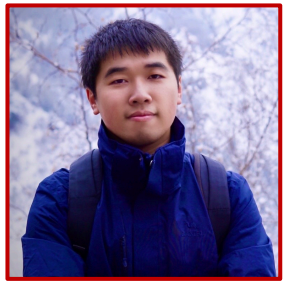
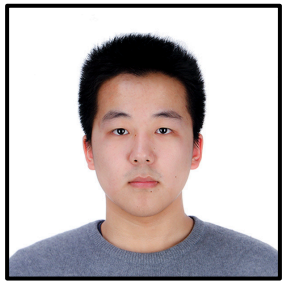
Unsupervised Cross-Task Generalization via Retrieval Augmentation




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

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

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

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

NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2022

Massively Multi-Task NLP Models

Input: Is this review positive or negative? Review: *Best cast ...*

Output: *Positive*

$\mathcal{T}_i$: amazon_plority

Input: Make a title for news: *British mathematician Stephen Haw...*

Output: *Mathematician Hawking ...*

$\mathcal{T}_j$: gigawords

upstream tasks

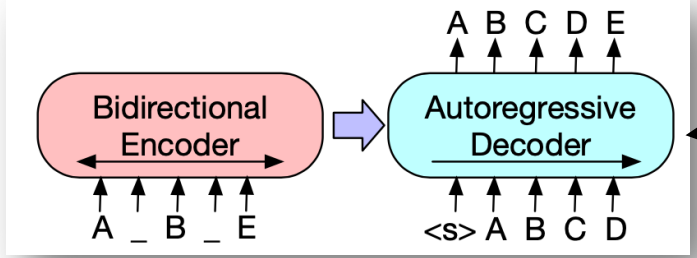
D

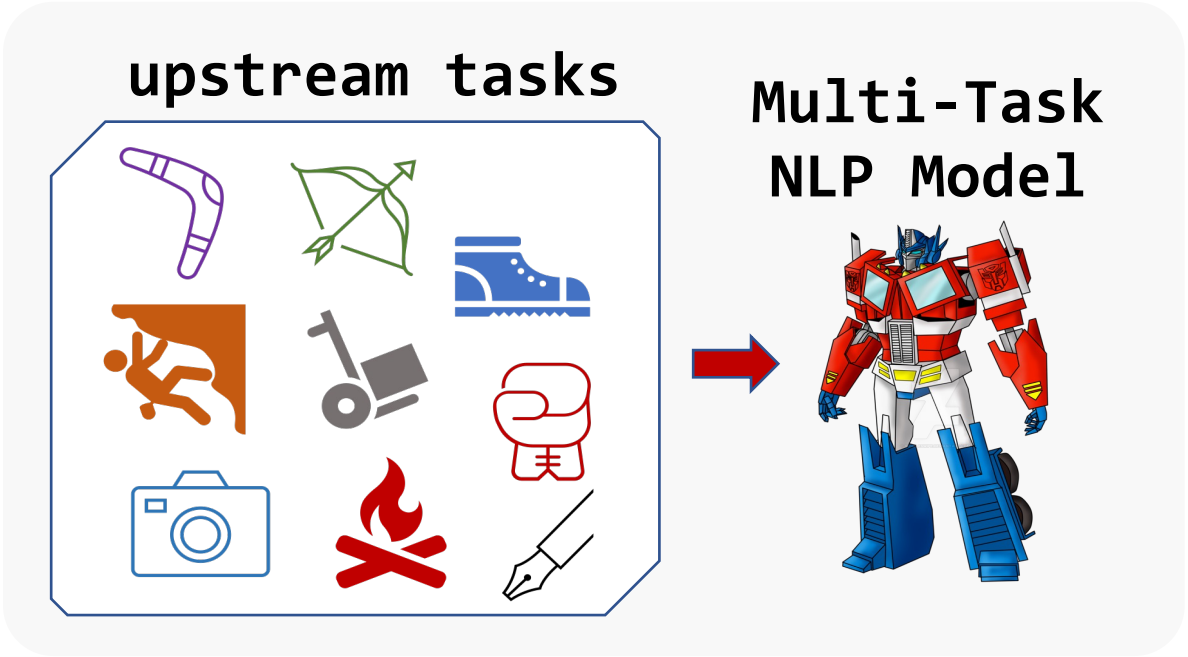
$\{\mathcal{T}_1, \ldots, \mathcal{T}_N\}$

Upstream Training

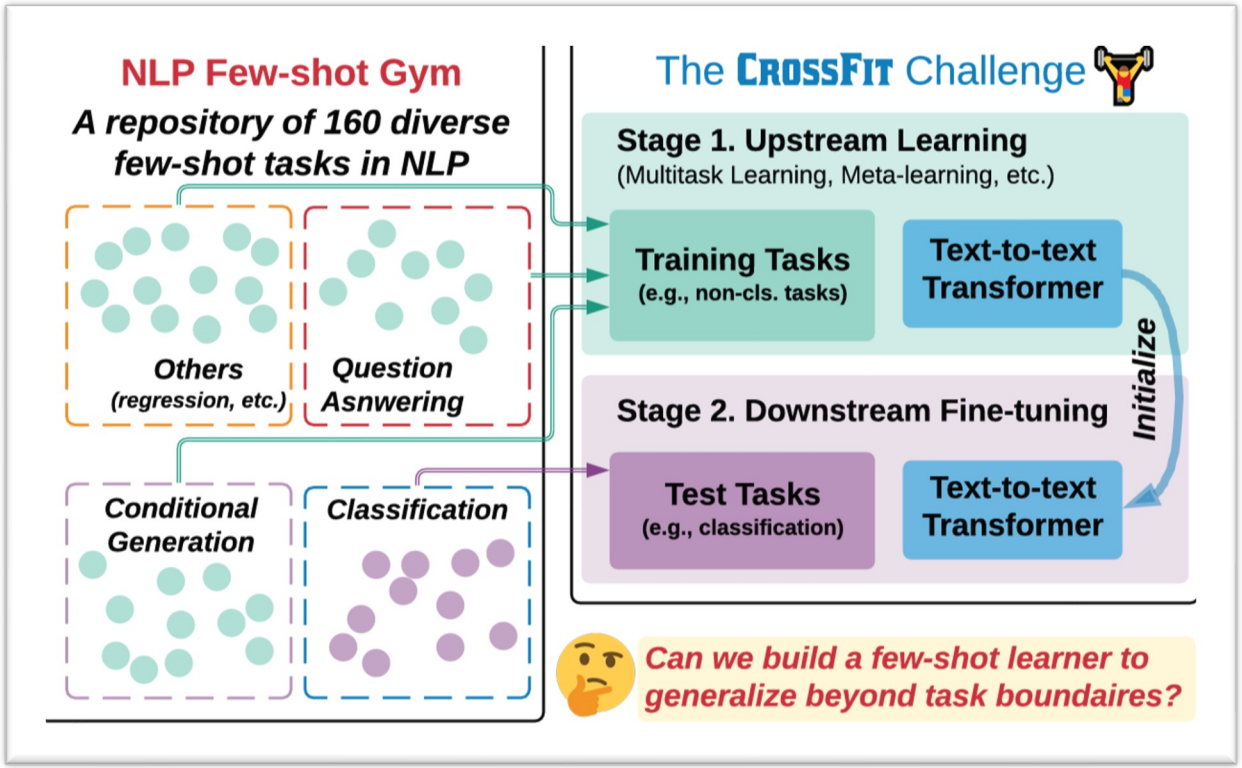
$\mathcal{M}$

Multi-Task NLP Model

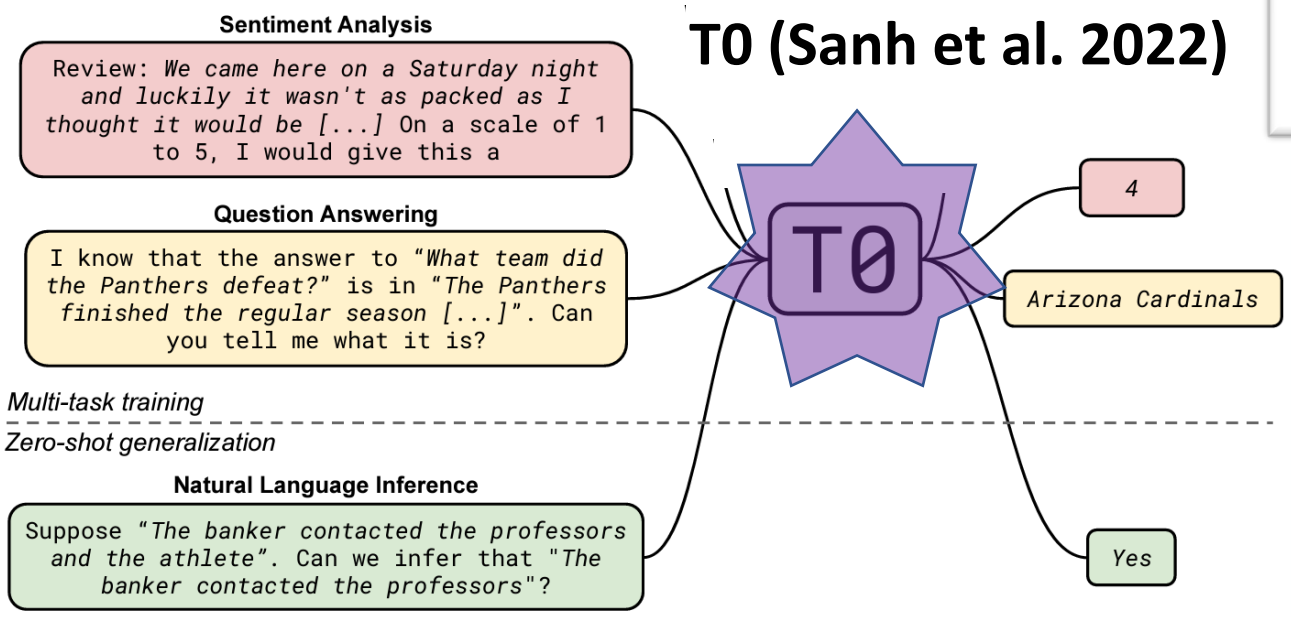




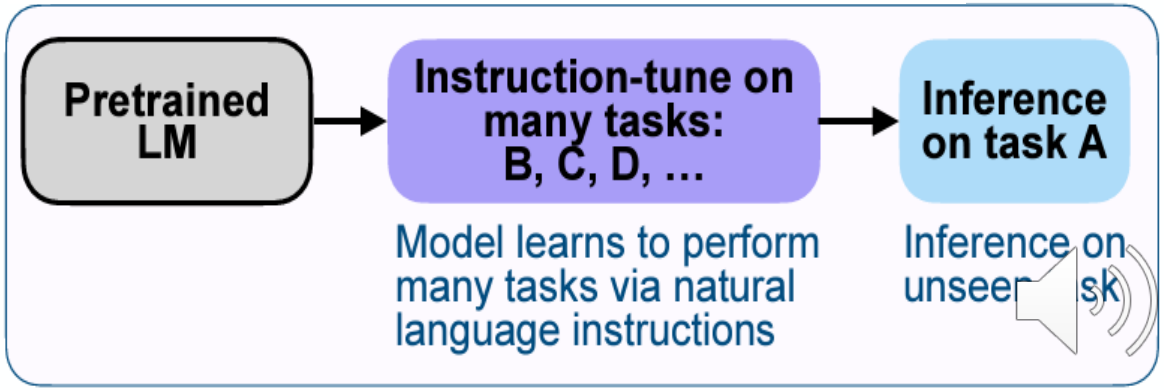
CrossFit (Ye et al. 2021)



T0 (Sanh et al. 2022)

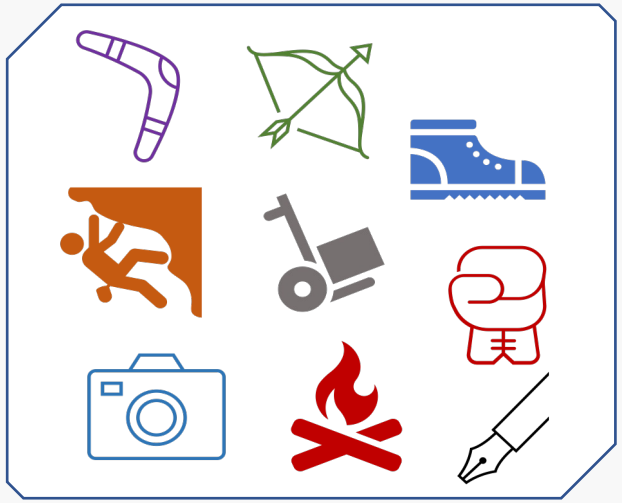


FLAN (Wei et al. 2022)

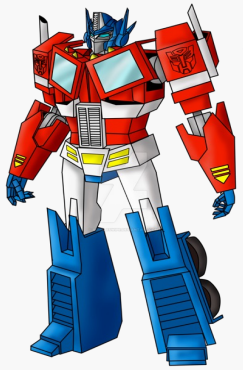


Unsupervised Cross-Task Generalization

upstream tasks



Multi-Task NLP Model



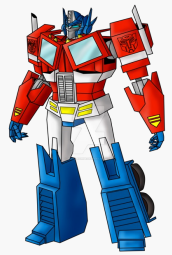
I want to ski!

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

Q_i

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

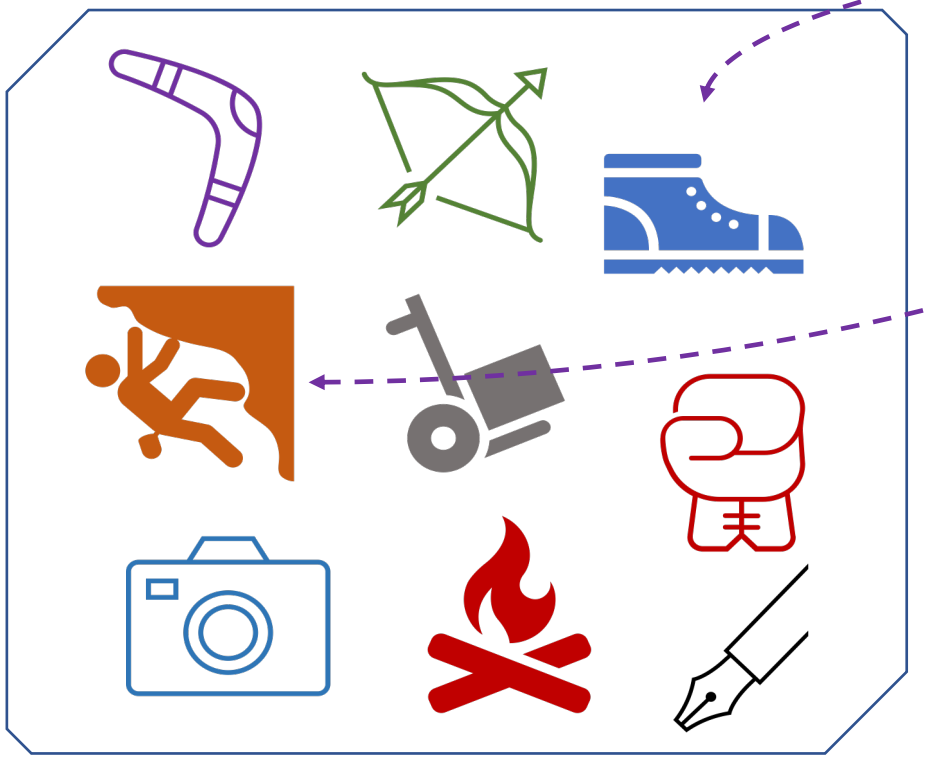
Input: Does "her" refer to Alice or Mary in this sentence: Alice ... Mary... and her ...



Unsupervised Cross-Task Generalization

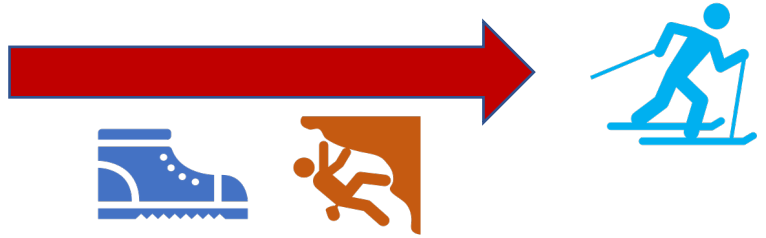
Retrieval Augmentation

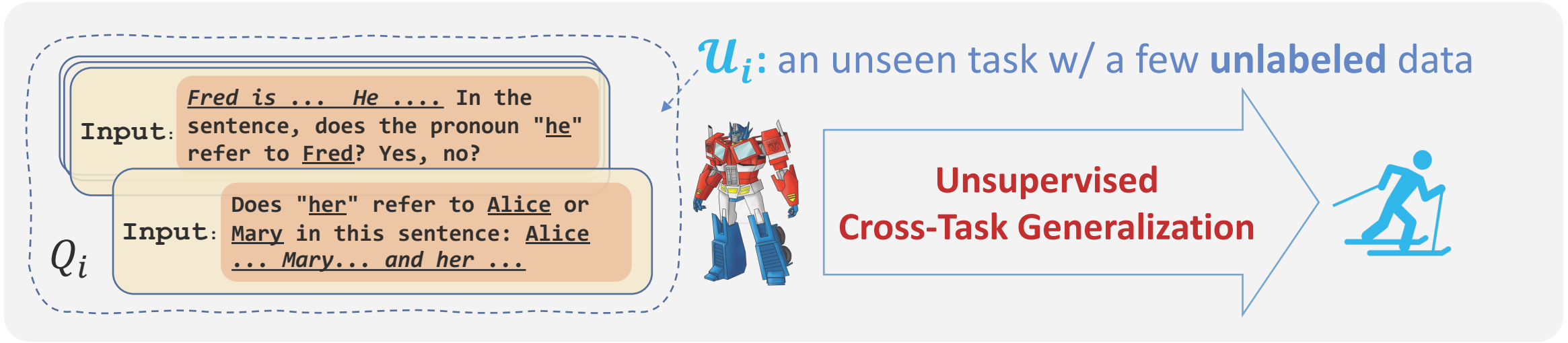
upstream tasks



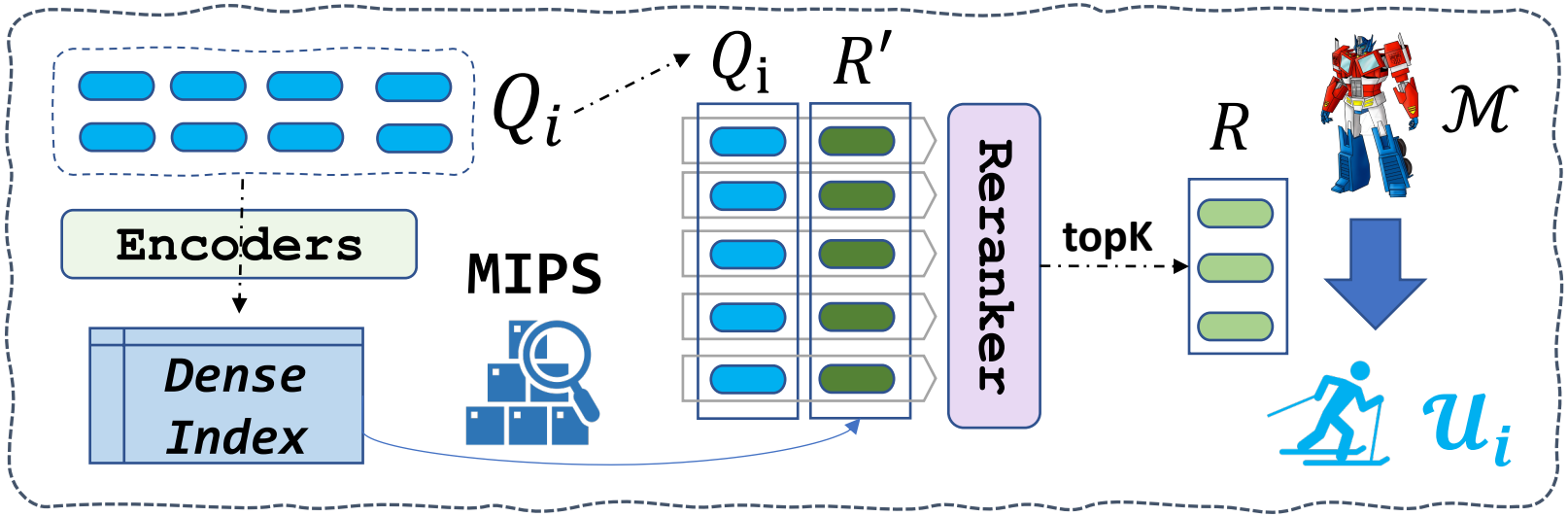
$\mathcal{U}_i$: an **unseen** task w/ a few unlabeled data

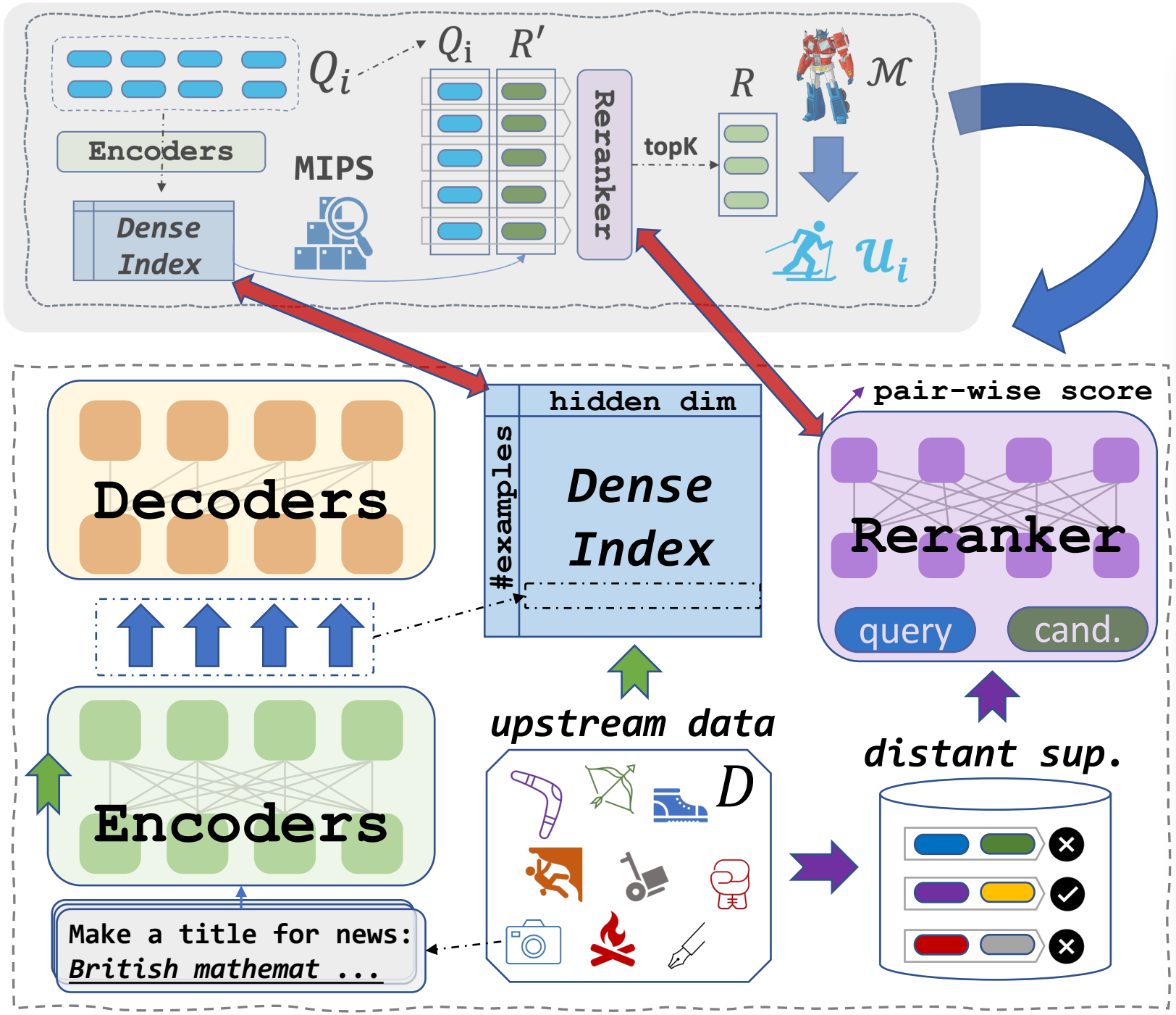
Retrieve some additional data for "**re-learning**" relevant skills.





The ReCross Framework





Algorithm 1: Distant Supervision Creation

Input: $\mathcal{M}$; D; $\mathcal{T}_q$
Output: $Z = (Z_q, Z_p, Z_n)$

$D_{\mathcal{T}_q} \leftarrow \{x \in D | x \text{ is an example of } \mathcal{T}_q\}$
$Z_q \leftarrow \text{Sample}(D_{\mathcal{T}_q})$; $H_q \leftarrow \text{Sample}(D_{\mathcal{T}_q})$
$R_Z \leftarrow \text{DenseRetrieve}(Z_q, D)$
/* 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 */
  $\{G_1, ..., G_n\} \leftarrow R_Z.\text{split}()$
  foreach G_i *in* $\{G_1, ..., G_n\}$ **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**
      $scores[x].\text{append}(\ell)$ /* Score each example in the group w/ the loss. */
/* Use mean group score as score for single examples */
foreach $x \in R_Z$ **do**
  $score[x] \leftarrow \text{mean}(scores[x])$
/* Sort R_Z by $score$ in increasing order. */
$R_Z.\text{sort}(\text{key: score, order: increasing})$
$Z_p \leftarrow$ First W items of R_Z
$Z_n \leftarrow$ Last W 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/**