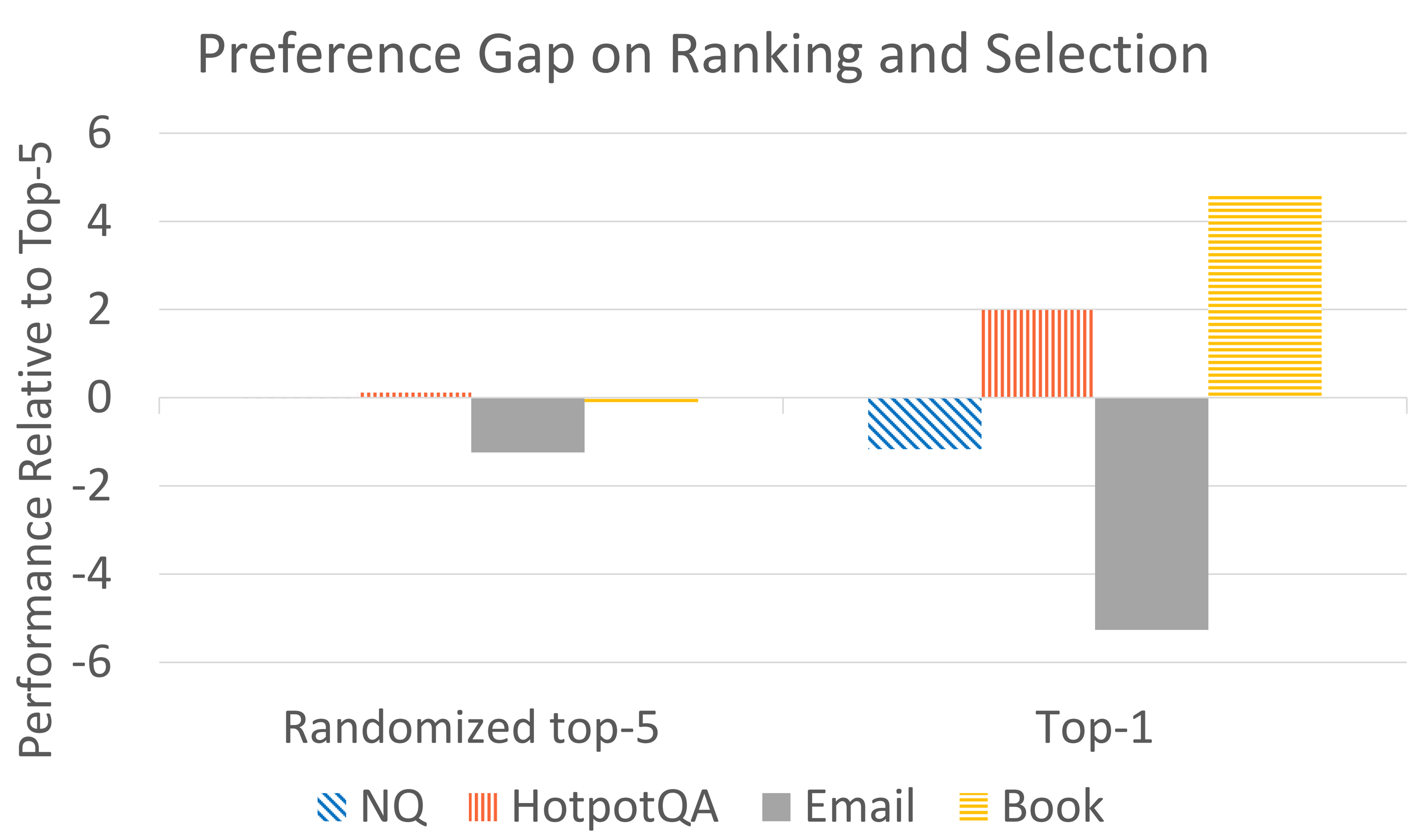
Bridging the Preference Gap between Retrievers and LLMs

Zixuan Ke[1], Weize Kong[2], Cheng Li[2], Mingyang Zhang[2], Qiaozhu Mei[3] and Micheal Bendersky[2]
University of Illinois at Chicago[1], Google Research[2], University of Michigan [3]

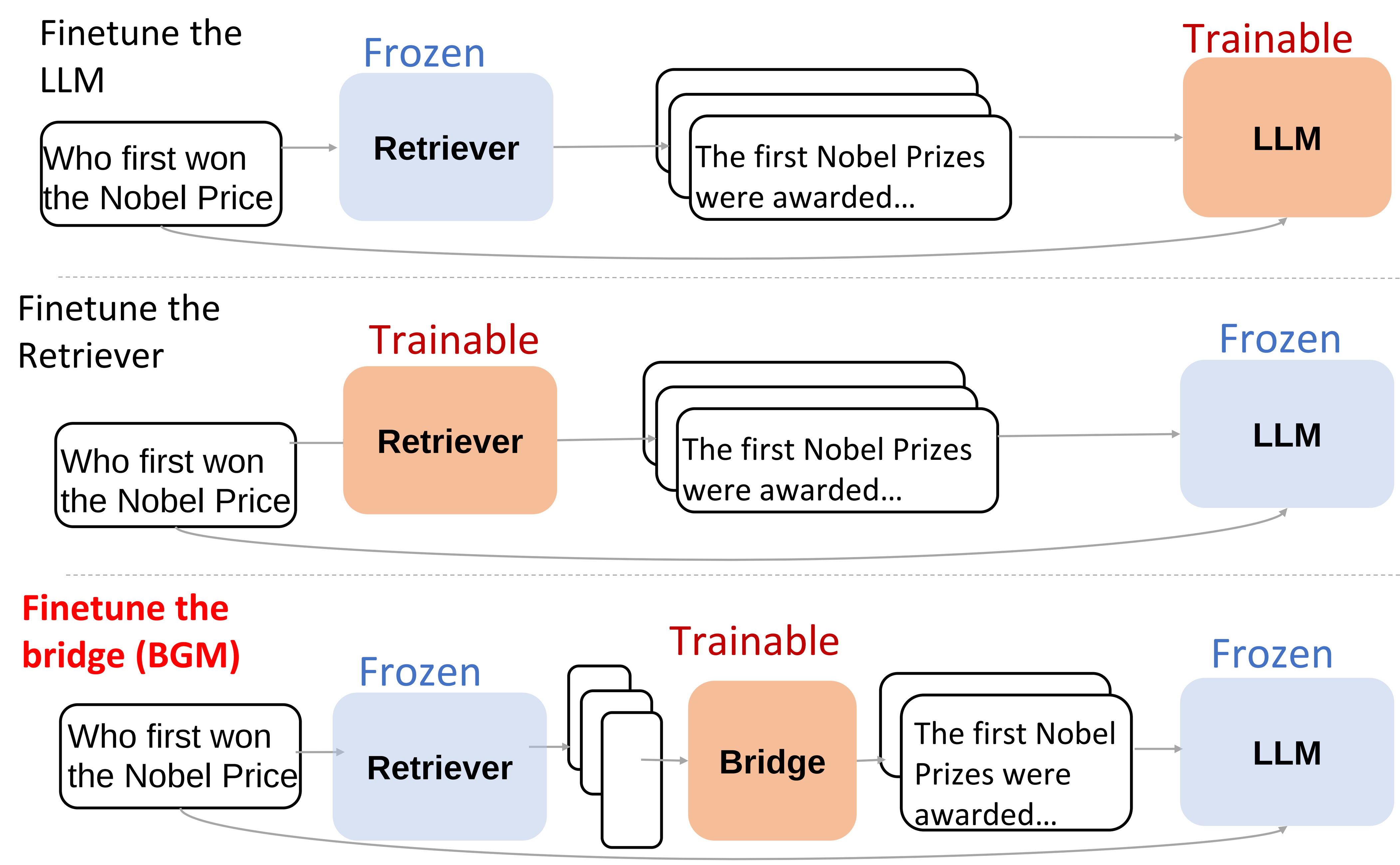
- **Existing** RAG studies retrievers and LLMs **separately**
  - **Retriever**: ranking is **the most important** as human read from top to bottom
  - **LLM**: exhibits preferences **different** from humans!
  - **There is a preference gap!**
- **Our goals:**
  - Investigate the existence of the preference gap
  - Bridge the gap via a seq2seq bridge model
  - Evaluate BGM with diverse tasks: QA and text generation, publicly available and personalized datasets.

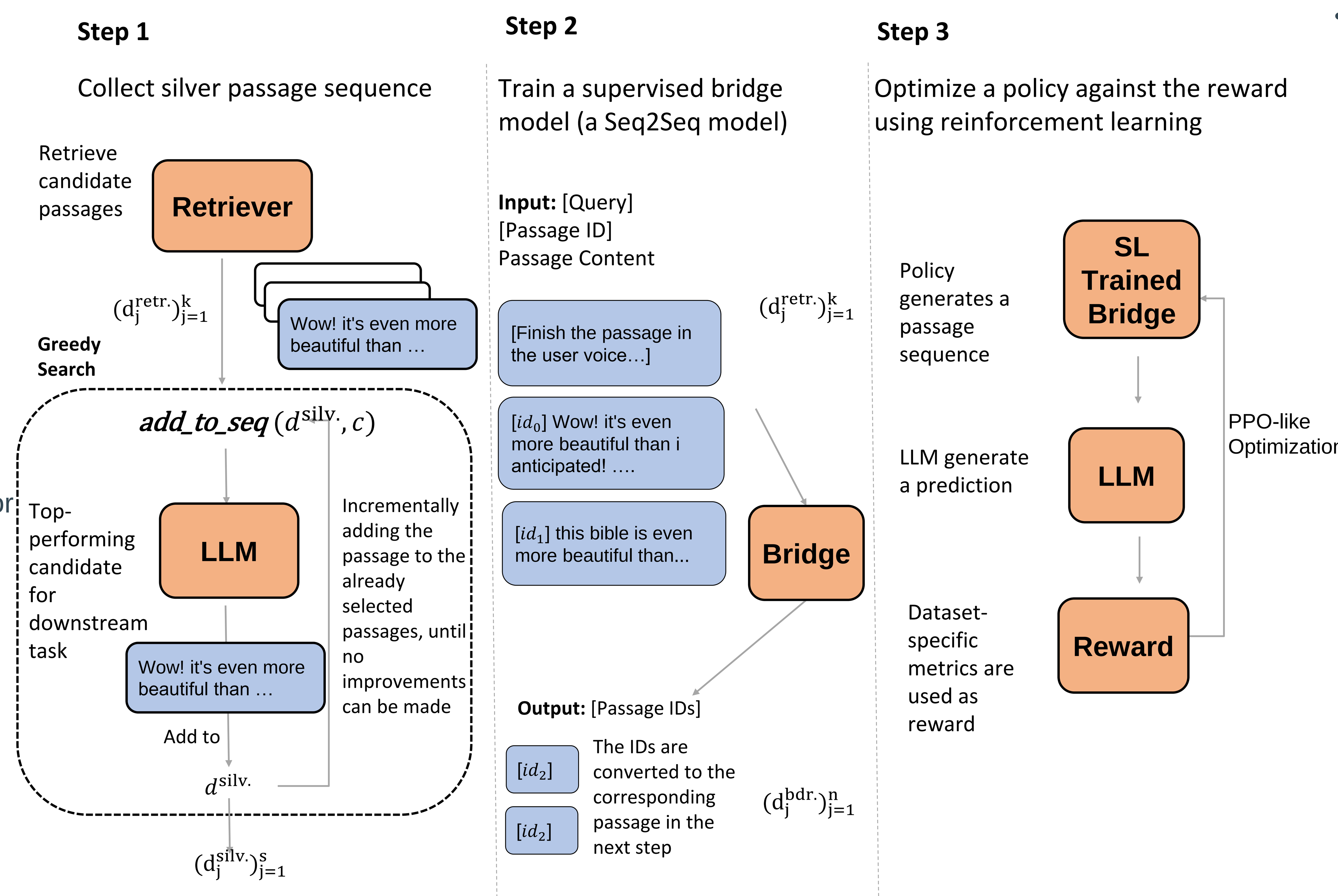
Preference Gap



- **Preliminary Experiments:**
  - Retrieve passages using off-the-shelf retriever
  - Use top K of the retrieved passages as additional context for a frozen Palm2-S LLM.
- **Key Observations:**
  - Randomizing the ordering of top-5 retrieved items (passages), the performance of RAG only varies by around 1%
  - The variation exceed 5% when the LLM is only presented with the top-1 passages under each order
- **Lesson:**
  - The general belief in ranking **does not** apply to LLMs!
  - We need to **bridge the preference gap**!

Bridge the Preference Gap





- **Bridge Model:**
  - Unlike previous RAGs that update LLMs, retrievers, or both. BGM connects a **frozen** LLM and a **frozen** retriever through a **seq2seq** bridge model
  - **Input:** Top-K passages given by a **frozen** retriever
  - **Output:** a sequence of passages (the sequence length is **varied,** due to **selection** or **repetition**)
  - **Goals:**
    - Adapt the retrieved information to the LLM's preference.
    - Not only **ranking,** but also **selection** and potentially **repetition**
- **Train the Bridge Model**
  - **Step 1: Supervised Learning**
    - Data Synthesis: we obtain the silver passage sequence via **greedy search**
      - iteratively add the next best candidate passage to the sequence and measure the results based on the resulted task performance. Stop until no improvement can be made
  - **Step 2: Reinforcement Learning**
    - Reward: Performance of downstream tasks
    - Policy model: bridge model
    - Action space: passage IDs

Model Metric	NQ EM	HotpotQA EM	Email BLEU	Book BLEU
Naive	33.07	28.01	5.57	11.5
Random	43.71	26.10	8.55	8.61
GTR	43.79	25.80	9.76	8.75
PSR	43.60	25.51	9.08	9.14
BGM	**45.37**	**35.64**	**10.42**	**12.07**

Table 2: Performance of all 4 datasets.

✓ **BGM achieves the highest**
✓ Book dataset has less improved: retrieval is not always essential! Naïve can outperform other baselines as well
✓ **GTR < BGM:** Similar results given by GTR and BGM in NQ: most instances in NQ need only one passage
✓ **PSR < BGM:** Ranking (PSR) is not sufficient. **Selection must be taken into account**
✓ (in the paper) Naïve manual threshold applied to reranking model is also insufficient

SCAN ME