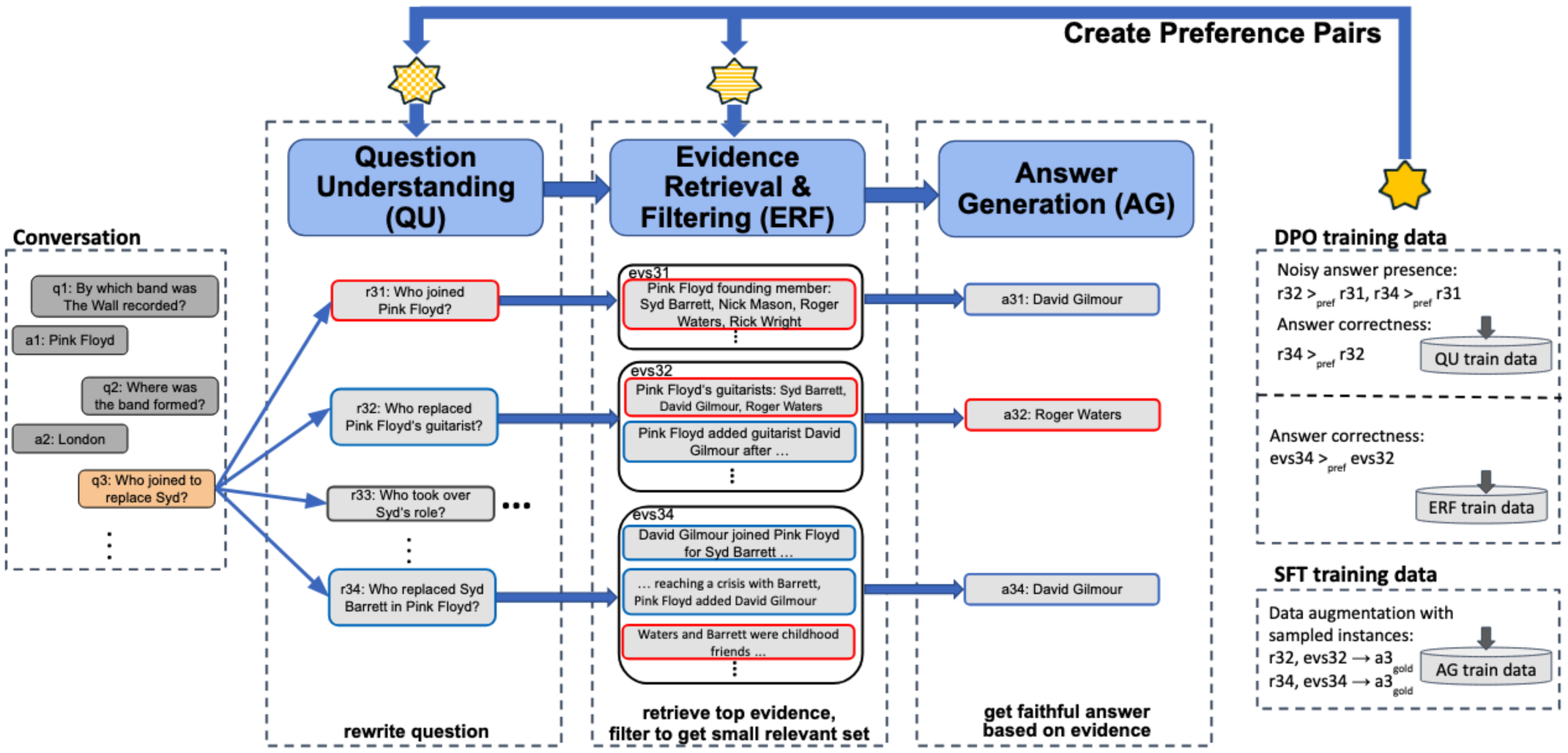
Preference-based Learning with Retrieval Augmented Generation for Conversational Question Answering

Magdalena Kaiser and Gerhard Weikum
Max Planck Institute for Informatics, Saarland Informatics Campus, Germany
mkaiser@mpi-inf.mpg.de, weikum@mpi-inf.mpg.de



PRAISE PIPELINE FOR CONVQA

- **QU:** question reformulations based on conv. history, most beneficial for retrieval and answer generation
- **ERF:** initial retrieval step + judiciously designed LLM-based evidence filtering technique
- **AG:** answers based on rewritten question + evidence
- Trained **LLM adapters** for each subtask

CHALLENGES FOR TRAINING

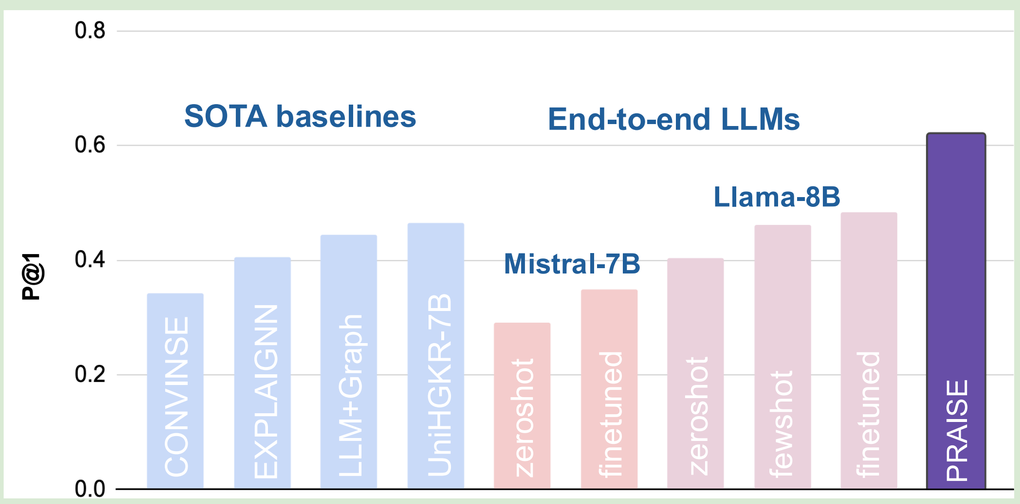
- **No intermediate supervision data** (for QU and ERF)
- Relying on **human feedback** is **expensive**

PRAISE TRAINING

1. **Create samples** using **initial pipeline** (few-shot QU, retrieval-only ERF, AG with benchmark data)
2. **Evaluate** answering **performance** for samples
3. **Create preference data** for DPO training:
   - **For QU:** samples preferred where answer in evidence AND final answer correct
   - **For ERF:** evidence set preferred that leads to correct answer
4. Additional **SFT training data for AG** based on generated samples

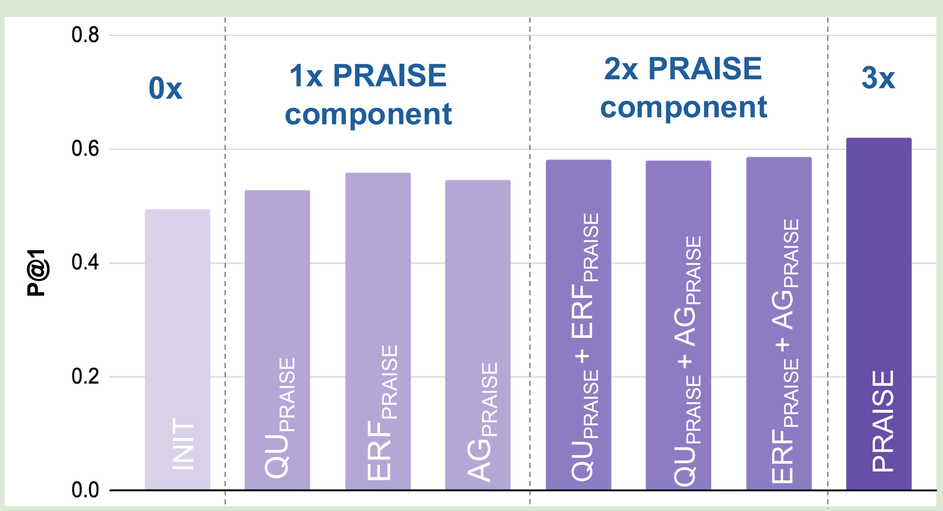
EXPERIMENTAL RESULTS

- On **ConvMix benchmark** (over KG + text, 5 domains)
- PRAISE achieves **new SOTA performance** (+15.5%)
- **Pipeline-based better** than **end-to-end** LLMs



Main results (Precision@1) comparing PRAISE to end-to-end LLMs and other competitors on ConvMix

- **All pipeline stages contribute:** highest performance with all three components
- PRAISE retains **high answer presence** in top-50:
  - **Initial:** AP@500=73% → AP@50=49%
  - **PRAISE:** AP@500=77% → AP@50=76%



Effect of different pipeline components in PRAISE (Precision@1 on ConvMix)

MAX PLANCK INSTITUTE FOR INFORMATICS

SIC Saarland Informatics Campus