

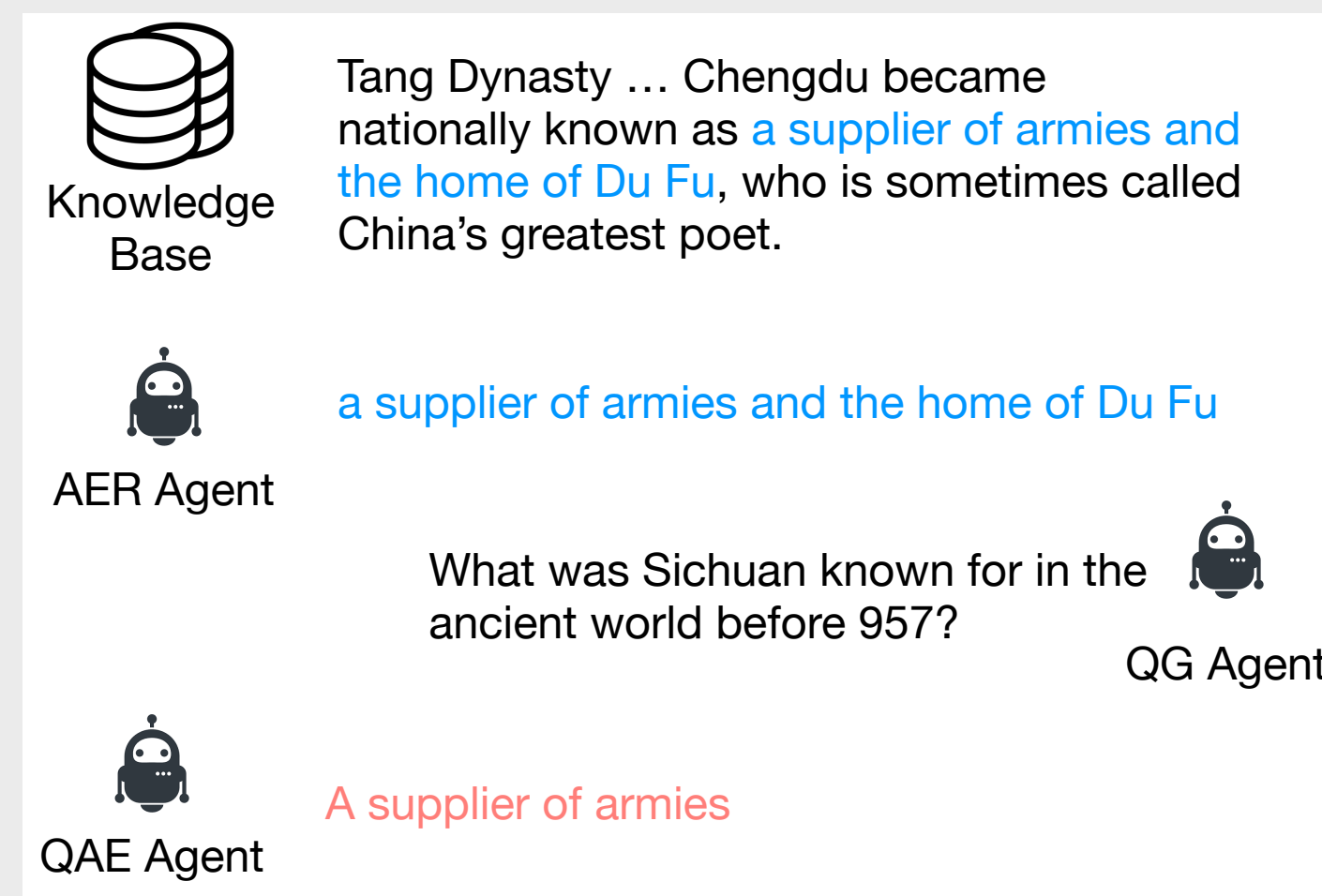
Cooperative Self-Training for QA



Hongyin Luo (hyluo@mit.edu), Shang-Wen Li, Mingye Gao, Seunghak Yu, James Glass (glass@csail.mit.edu)

Abstract

Generate synthetic QA for plain texts with **RGX**
 Ans. entity recognition (AER)
 Ques. generation (QG)
 Ans. extraction (AEX)

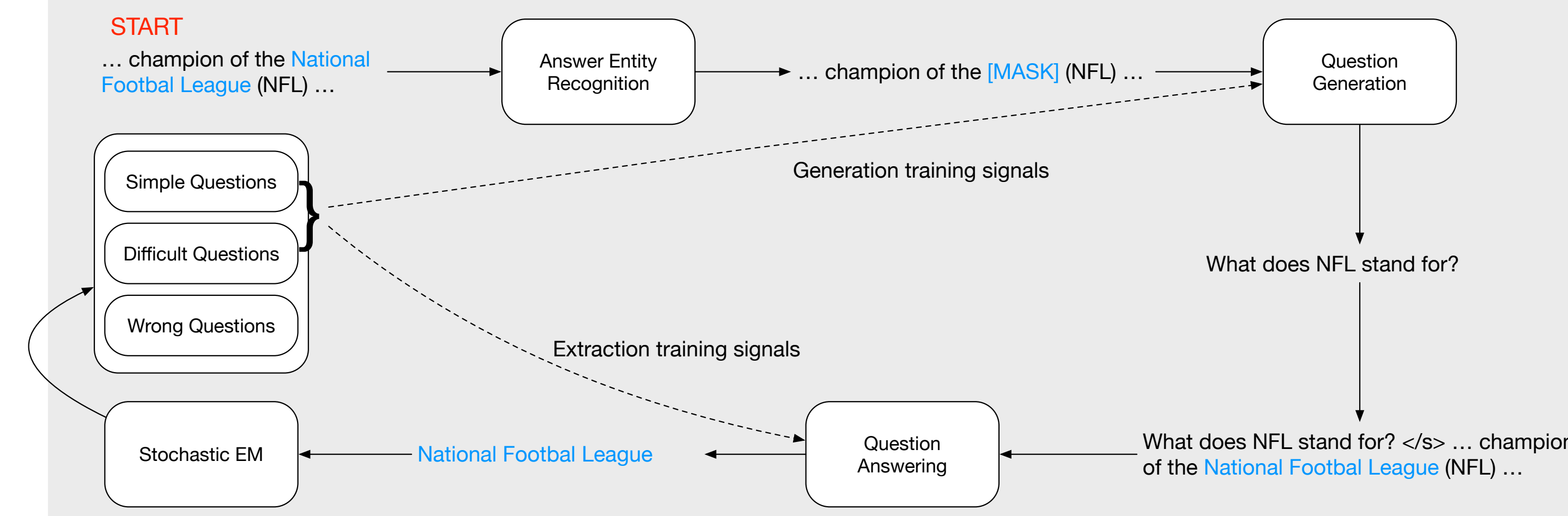


Cooperative self-training of question generation and answering models. Zero-shot adaptation to new domains

1. Generate synthetic QA pairs of the target domain
2. Self-training the QG model selected QA pairs
3. Generated new QA pairs with the updated QG model
4. Self-training the AEX model with selected QA pairs

Cooperative Self-Training Pipeline

1. Recognize potential answer entities (AE) w/ the AER model
2. Generate questions based on recognized AEs
3. Answer generated questions w/ the pretrained AEX model
4. Predict the quality of synthetic QA pairs
5. Tuning QG and AEX models with high-quality QAs



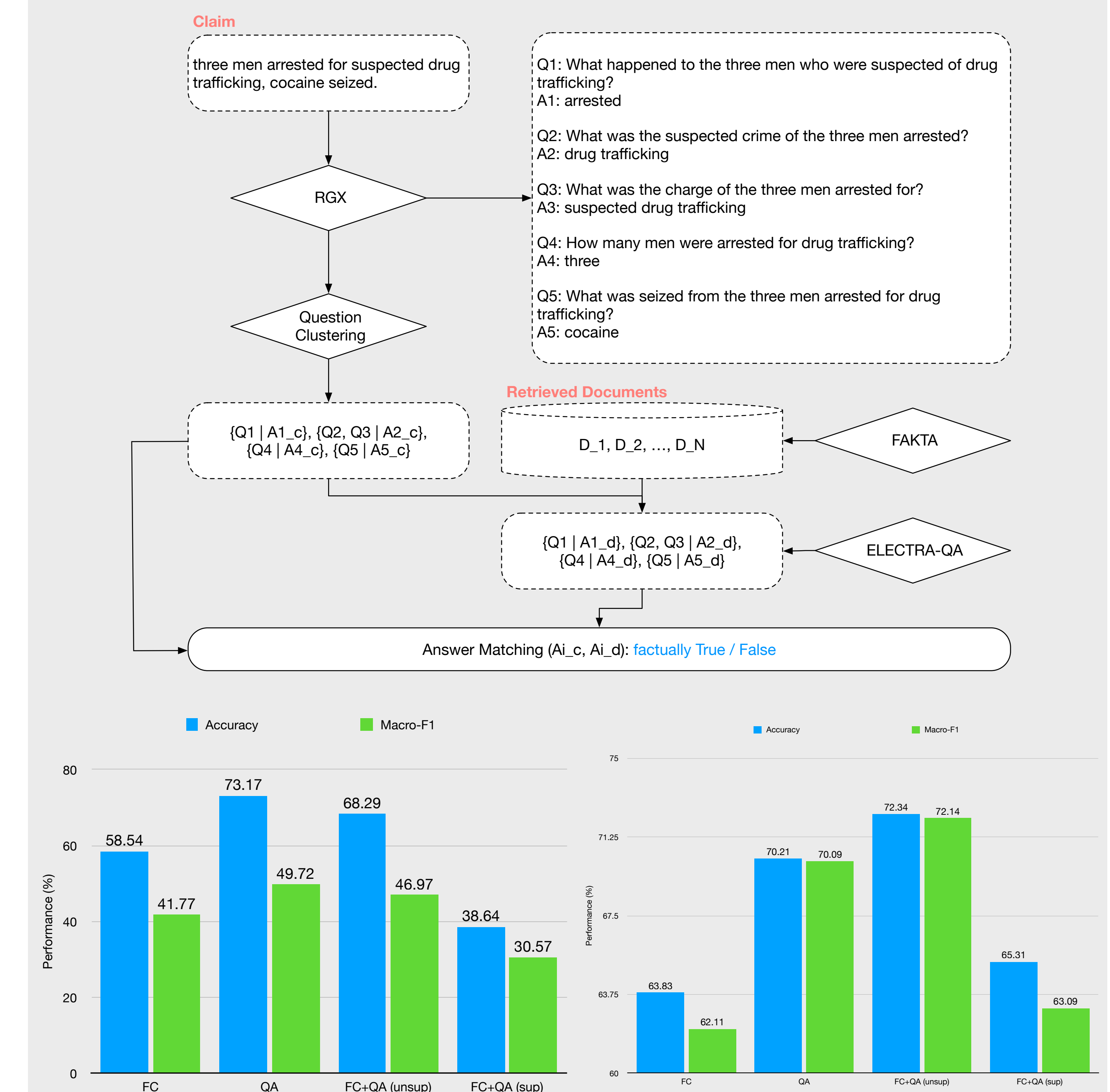
Experiment Results

Training on SQuAD v1.1, zero-shot adaptation on MRQA

Dataset	ELECTRA		QAGen2S		RGX + Coop. ST	
	EM	F1	EM	F1	EM	F1
BioASQ	58.7	73.1	56.8	71.7	60.3	74.8
Txtbk-QA	43.0	53.6	48.0	56.5	51.2	61.2
RACE	38.3	52.5	43.4	54.9	44.9	58.7
Rel. Ext	79.0	88.4	73.4	84.8	79.2	88.6
DuoRC	53.1	64.2	53.3	64.6	57.4	66.2
DROP	48.3	60.8	42.2	54.5	47.6	60.9
Average	53.4	65.4	52.8	64.5	56.8	68.4

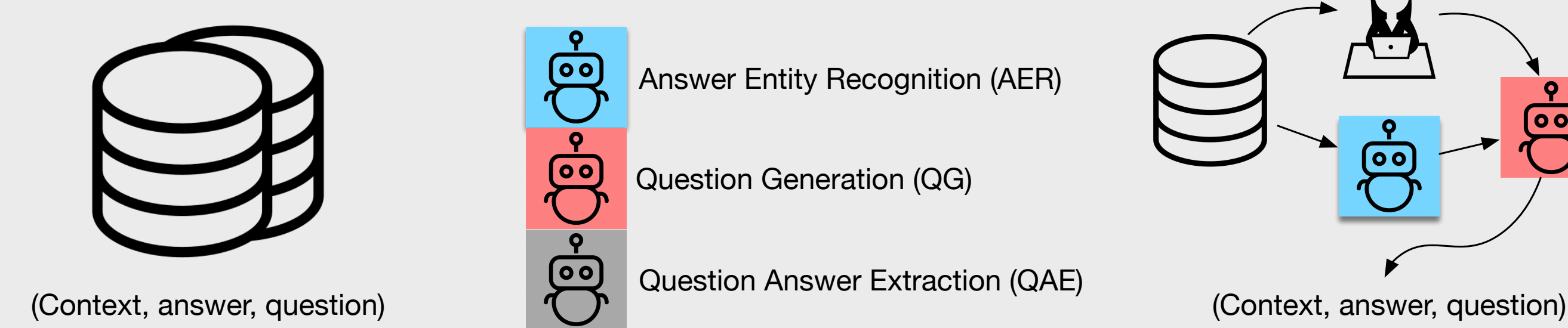
Future Directions

QA-based fact checking (FC)

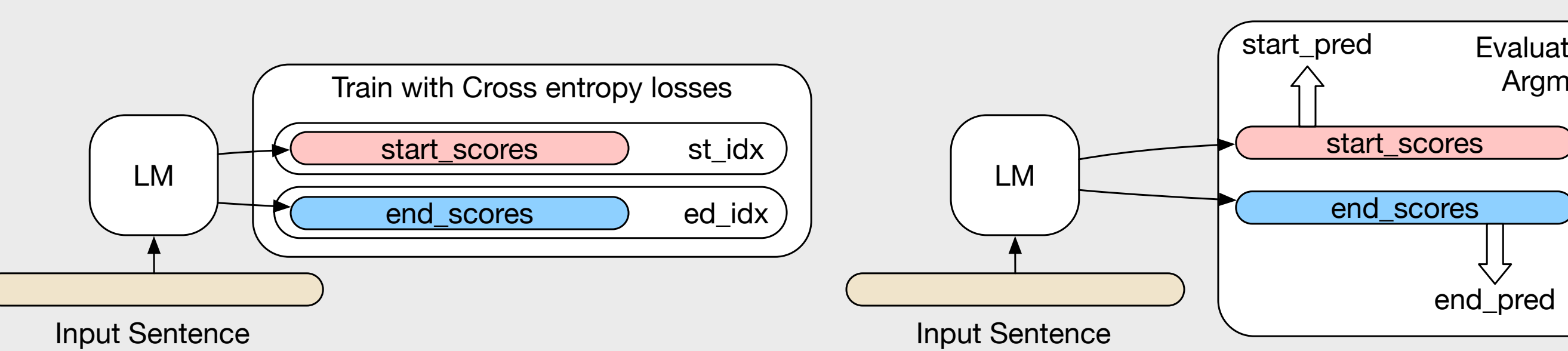
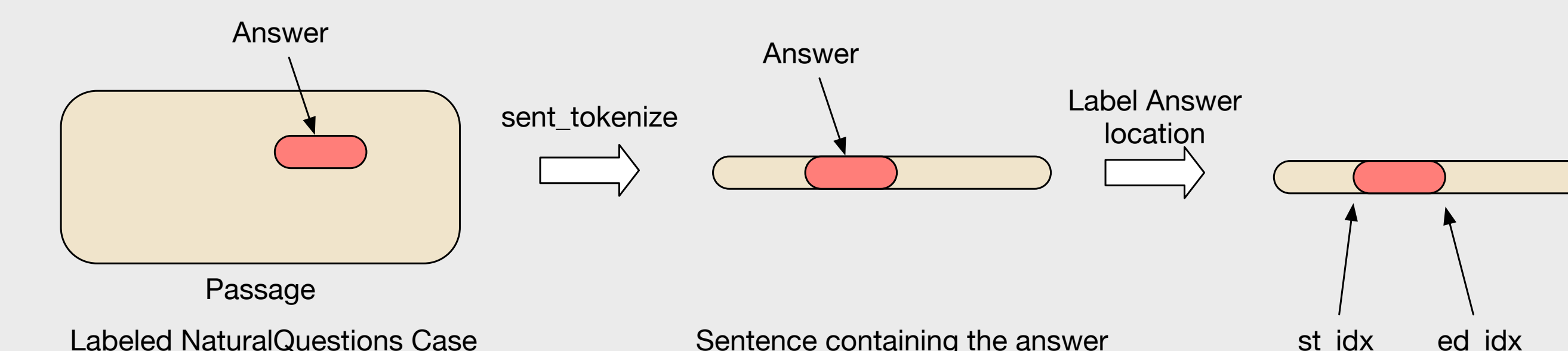


Synthetic QA Generation in RGX

1. Collect Seed Datasets
2. Pre-train Models
3. Adapt to New Domains



Answer entity recognition (AER)



Question generation (QG) and answer extraction (AEX)

Super Bowl 50 was an American football game to determine the champion of the [MASK] (NFL) for the 2015 season </s> **National Football League**

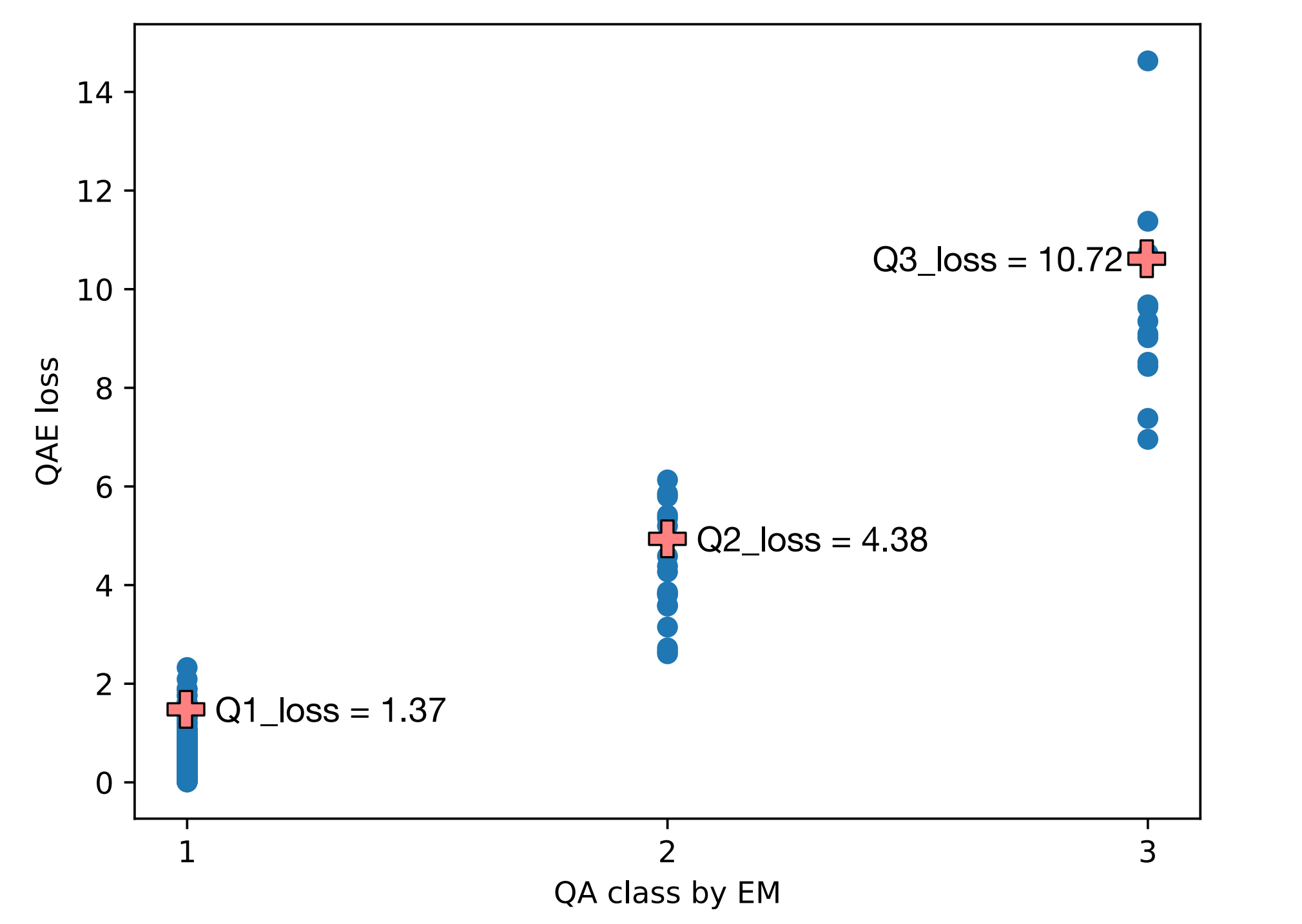
Recognized Answer Entity: National Football League
Generated Question: What does NFL stand for?
Predicted Answer: the National Football League

QA Quality Prediction with EM

Measure question quality by < QG loss + AEX loss >

Context: Despite differences in the spectrum of mutations in CN or CyN, **type or localization of mutation** only partially determine the clinical phenotype.

- Q1: What determines the clinical phenotype of a person with a mutation?
 Q2: What determines the clinical phenotype of a mutation?
 Q3: What is the only way to determine the clinical phenotype of a mutation?



Mutual Information QA + AER

$$\text{Log MI}(q, a, c) = \text{Log } P(a | q, c) + \text{Log } P(q | a, c)$$

Dataset	ELECTRA		Top-k + MI		AER + MI	
	EM	F1	EM	F1	EM	F1
BioASQ	58.7	73.1	57.8	72.9	59.9	74.0
Txtbk-QA	43.0	53.6	44.6	54.9	45.3	55.4
RACE	38.3	52.5	38.1	52.4	39.7	54.1
Rel. Ext	79.0	88.4	78.6	88.3	79.2	88.6
DuoRC	53.1	64.2	52.6	64.3	53.8	65.1
DROP	48.3	60.8	46.7	60.8	49.7	61.5

Other NLP Tasks

- Task-oriented pretraining on synthetic data
- Few-shot learning with synthetic data
 - Domain adaptation
 - Task adaptation
- Federated learning with synthetic data