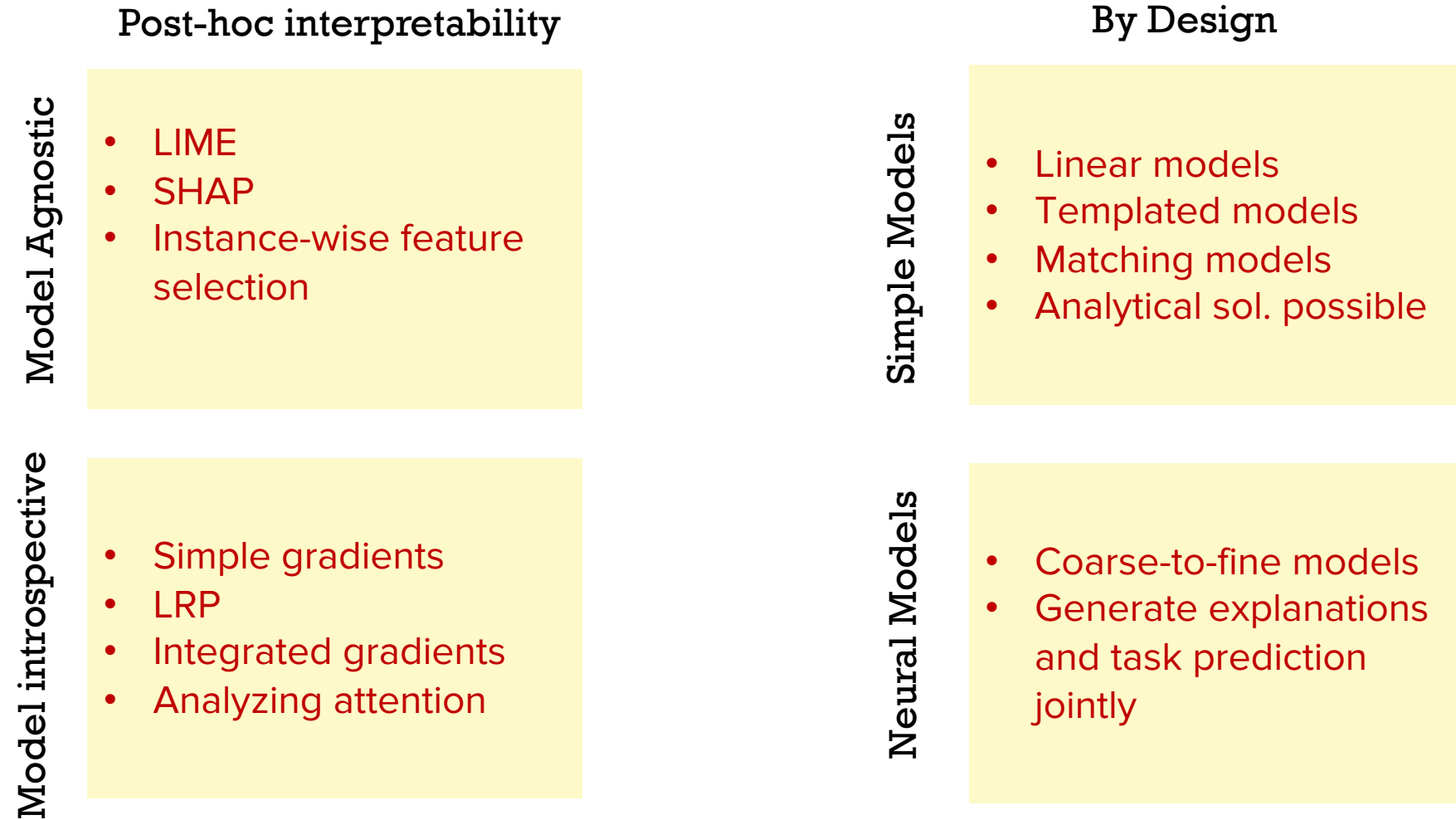# INTERPRETABILITY AND FEEDBACK

Interpretability Landscape

Coarse-to-fine Models

What is the capital of Australia ?

The country's other major metropolitan areas are Melbourne, Brisbane, Perth, and Adelaide. As the seat of the government of Australia, Canberra is home to many important institutions of the federal government, national monuments and museums. Canberra is also the capital of the country.

Select Sentences as Explanations
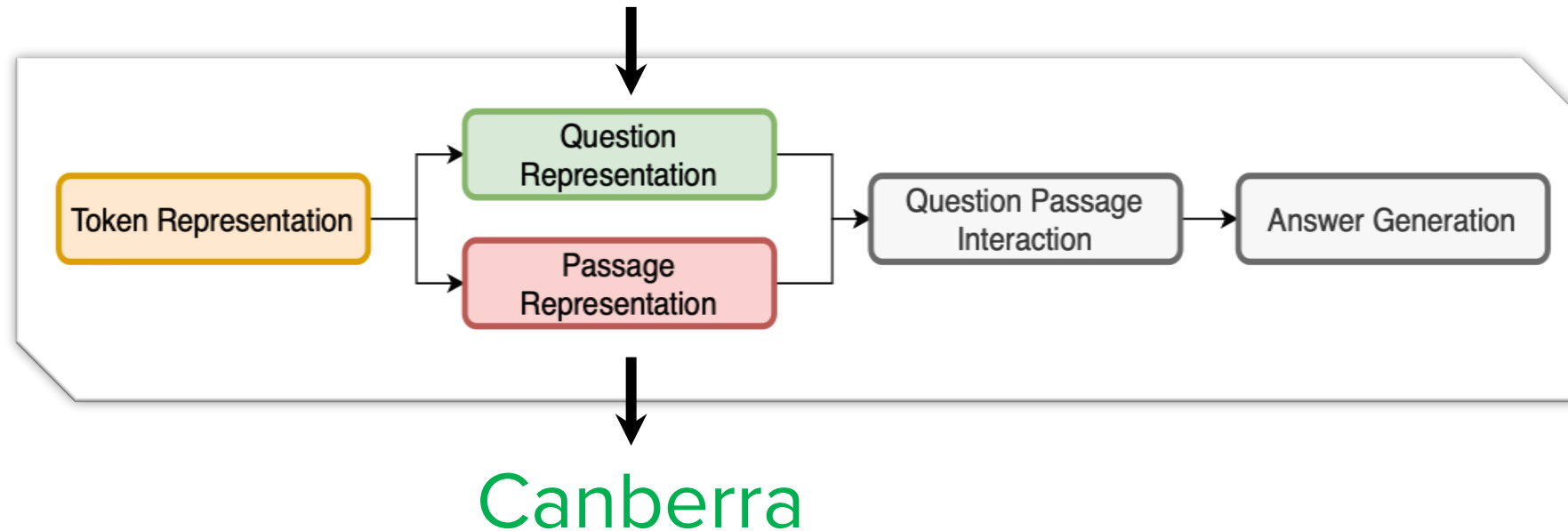
What is the capital of Australia ?

The country's other major metropolitan areas are Melbourne, Brisbane, Perth, and Adelaide. As the seat of the government of Australia, Canberra is home to many important institutions of the federal government, national monuments and museums.

Canberra is also the capital of the country.

Input to Reader

What is the capital of Australia ?

.... Canberra is also the capital of the country.

Selector network

Reader

Canberra

answer

Pipelined Models

- Sentence selection and answer predictions are independently trained

- What is the training data for sentence selection ?
  - Distance supervision
    - All sentences in the document containing answer is a positive instance
    - First sentence in the document containing the answer

- **Sentence selector** is trained on distantly supervised data

- **Answer predictor** is trained on the actual training data
  - Training data modified to only contain sentences selected from the selection stage
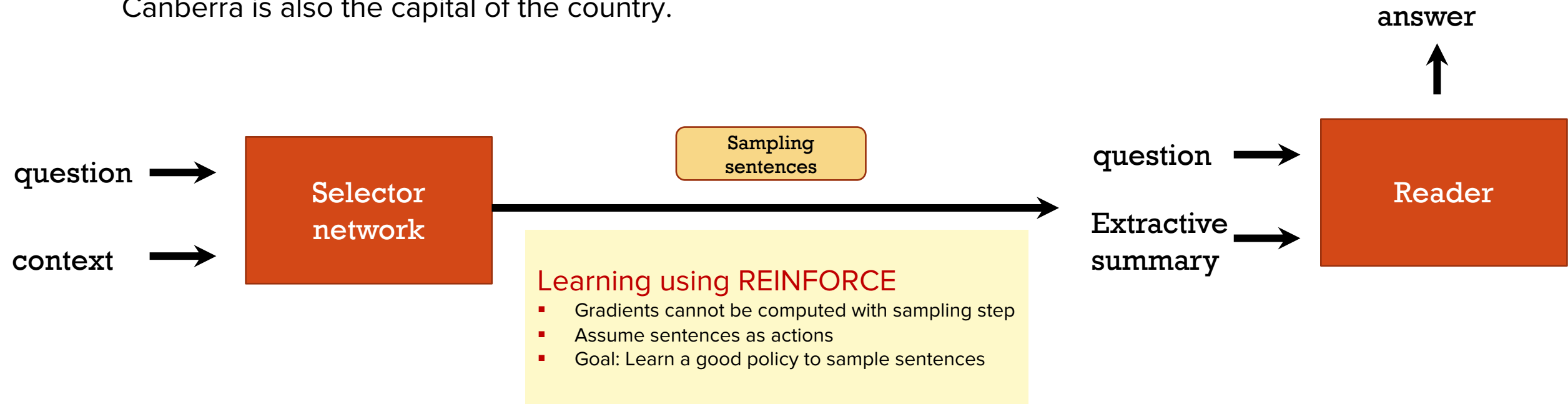
Selector network

↓

Reader

↓

answer

End-to-end Models

What is the capital of Australia ?

The country's other major metropolitan areas are Melbourne, Brisbane, Perth, and Adelaide.
As the seat of the government of Australia, Canberra is home to many important institutions of the federal government, national monuments and museums.
Canberra is also the capital of the country.

question →

context →

Selector network

Sampling sentences

question →

Extractive summary →

Reader

answer ↑

Learning using REINFORCE
- Gradients cannot be computed with sampling step
- Assume sentences as actions
- Goal: Learn a good policy to sample sentences

User Feedback

- Current systems assume a static collection, static training set

- In an online systems
  - Users continuously issue queries, provide implicit feedback

- How can we construct a continuously learning system from explicit user feedback ?
  - How do we use the feedback to update training set ?
  - Can we reconcile noisy and sometimes erroneous feedback ?

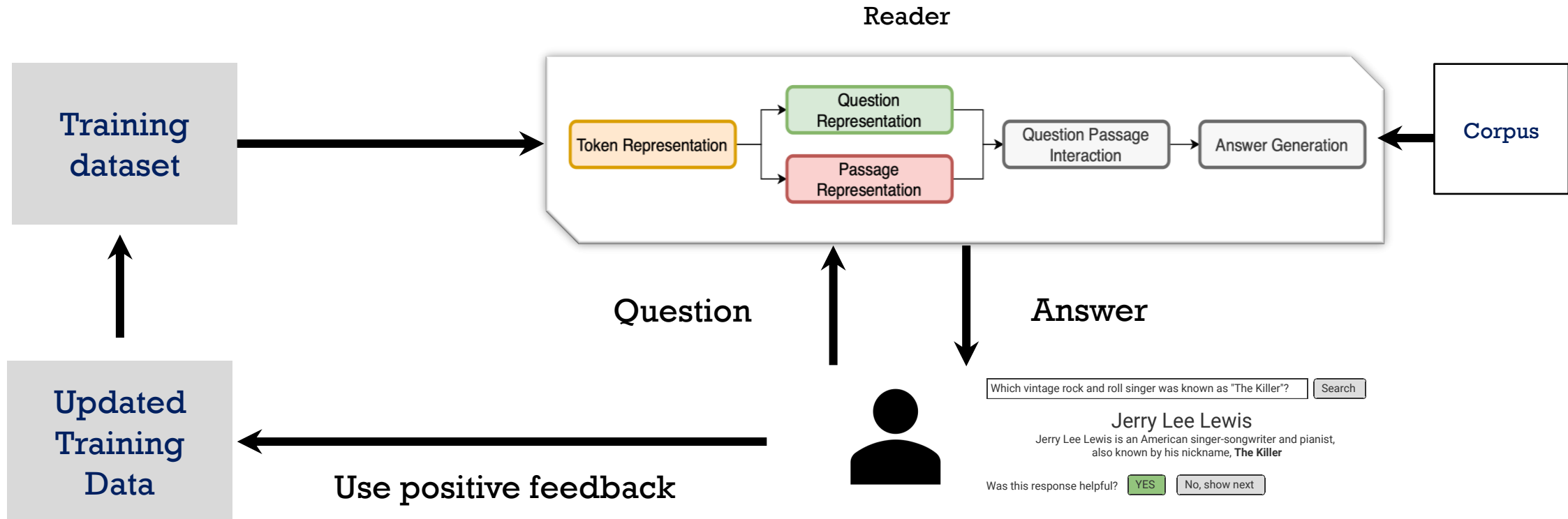| Which vintage rock and roll singer was known as "The Killer"? | Search |

Farrokh Bulsara

Farrokh Bulsara, known professionally as Freddie Mercury,
was a British singer, songwriter and record producer...

Was this response helpful?   YES   No, show next

Updating Training Set

Reader

Training dataset

Corpus

Question

Answer

Use positive feedback

Updated Training Data

Credibility Validation

Reader

Training dataset

Corpus

Question

Answer

Evaluate Feedback

Updated Training Data

Which vintage rock and roll singer was known as "The Killer"?   Search

Farrokh Bulsara, known professionally as Freddie Mercury, was a British singer, songwriter and record producer...

YES   No, show next

Credibility Validation
▪ Use question and answer using rest semantics to find support documents
▪ Supported by reader – top-k reader docs
▪ Length restrictions, NE restrictions

YES   No, show next

Was this response helpful?   YES   No, show next