

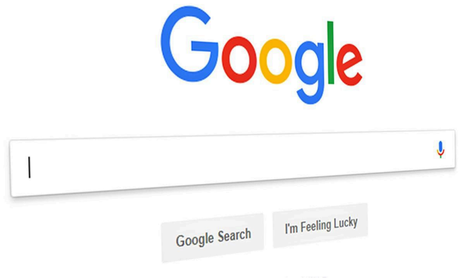

Zero-Shot Information Extraction as a Unified Text-to-Triple Translation

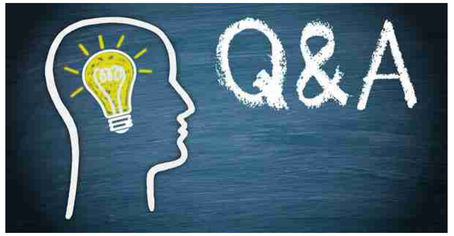
EMNLP 2021

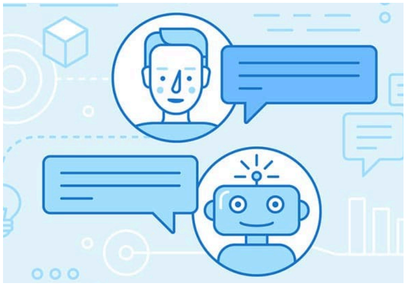
Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, Dawn Song

Information Extraction (IE): extract structures from unstructured data

NLP Applications







.....

Search Engine | Knowledge Base | QA System | Dialogue System

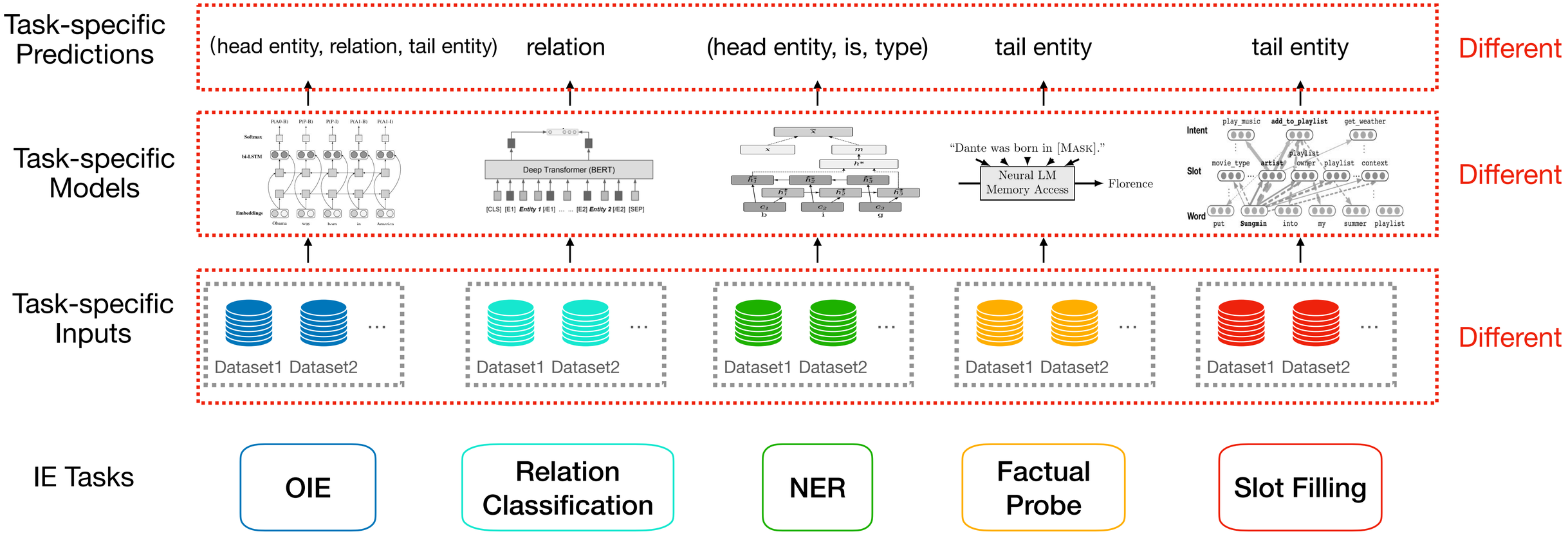
IE Tasks

- Open Information Extraction (OIE)
- Relation Classification
- Named Entity Recognition (NER)
- Factual Probe
- Slot Filling

.....

Information extraction is crucial to many NLP applications

We need a unified information extraction approach



There are many IE tasks, with different task-specific pipelines

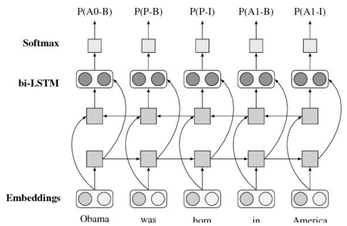
We need a unified information extraction approach

Task-specific Predictions

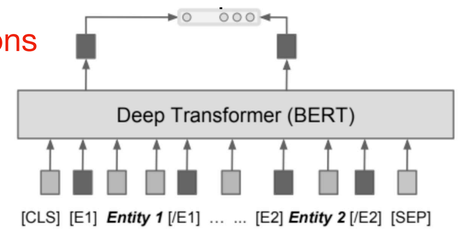
(Fisher, Born in; Glasgow)
(Fisher; is a graduate of; London Opera Centre)

city_of_birth

Task-specific Models



Different Task Formulations



Task-specific Inputs

Different Label Categories

Dataset1 Dataset2

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

Different Label Categories

Dataset1 Dataset2

Born in Glasgow, Fisher is a graduate of the London Opera Centre.
(Fisher, ?, Glasgow)

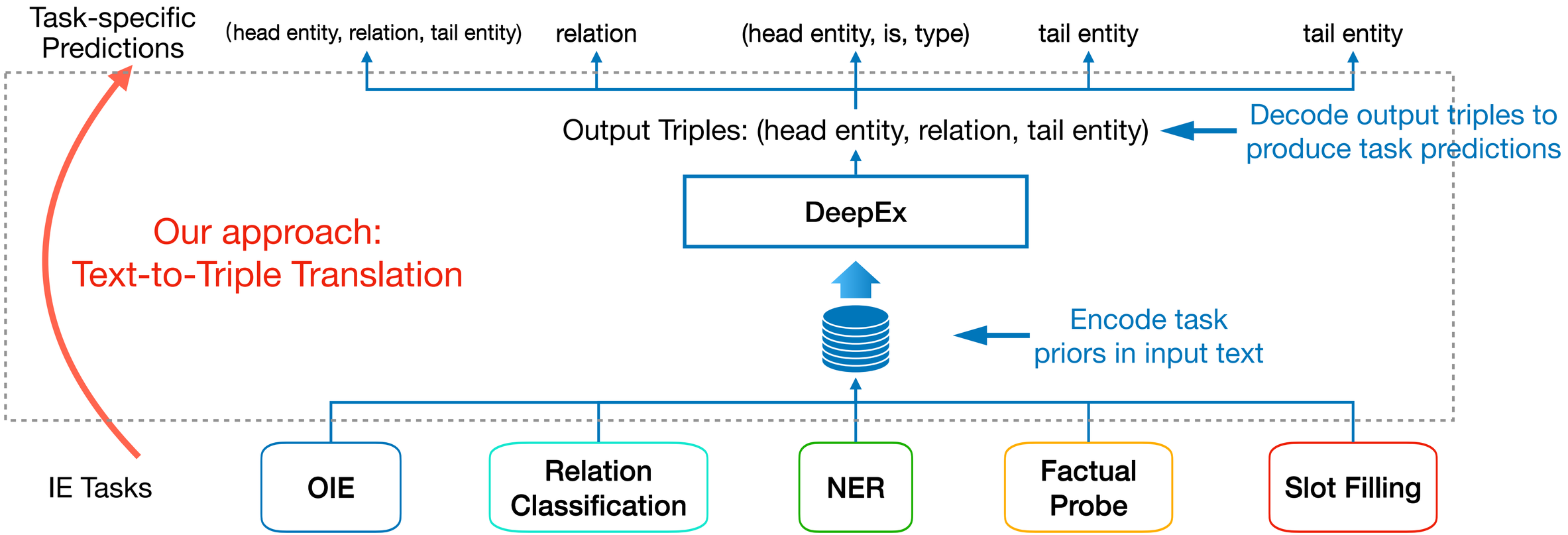
IE Tasks

OIE

Relation Classification

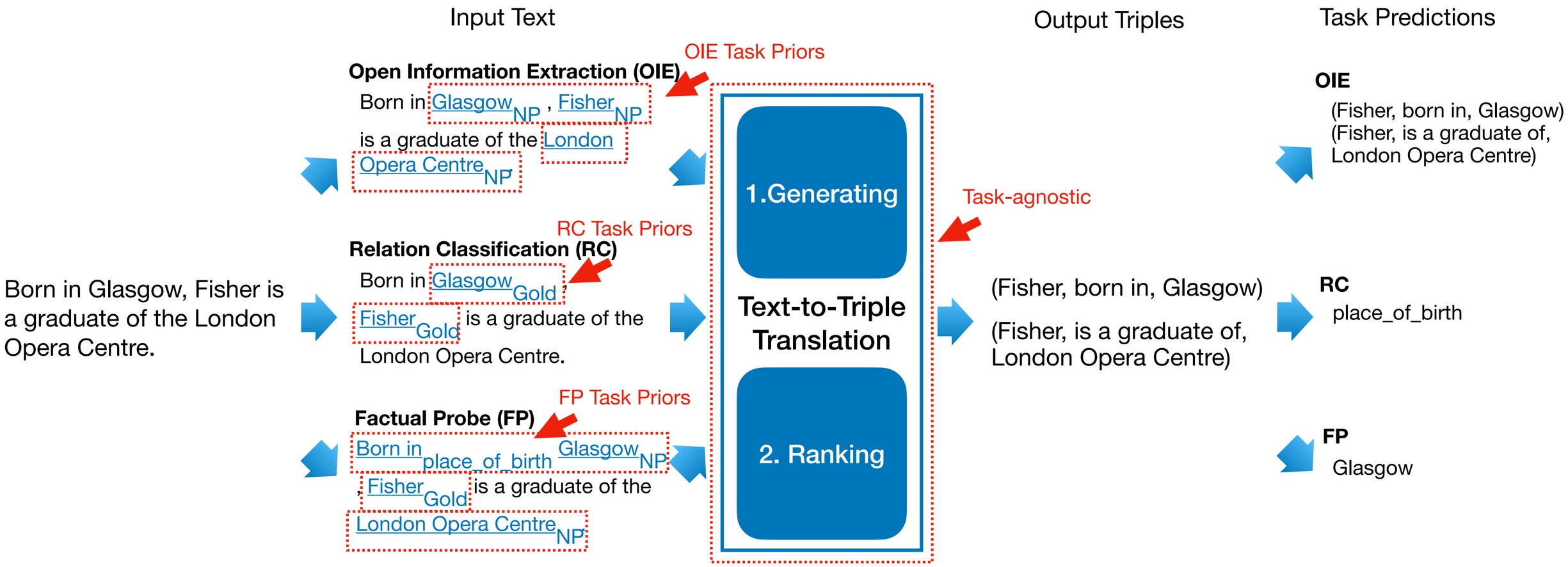
The main issue of existing IE methods: limited transferability

Our approach: a unified framework for information extraction



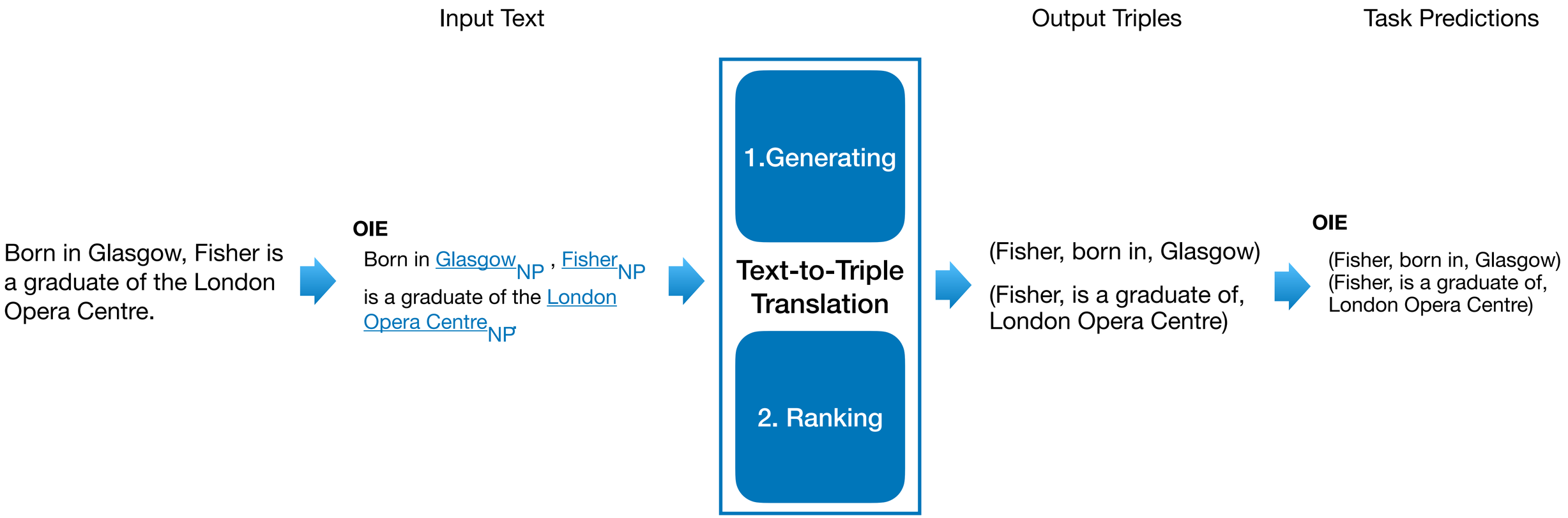
The basic idea: treat every information extraction problem as a “text-to-triple” problem, i.e., translating input text to output triples

Our method: text-to-triple translation

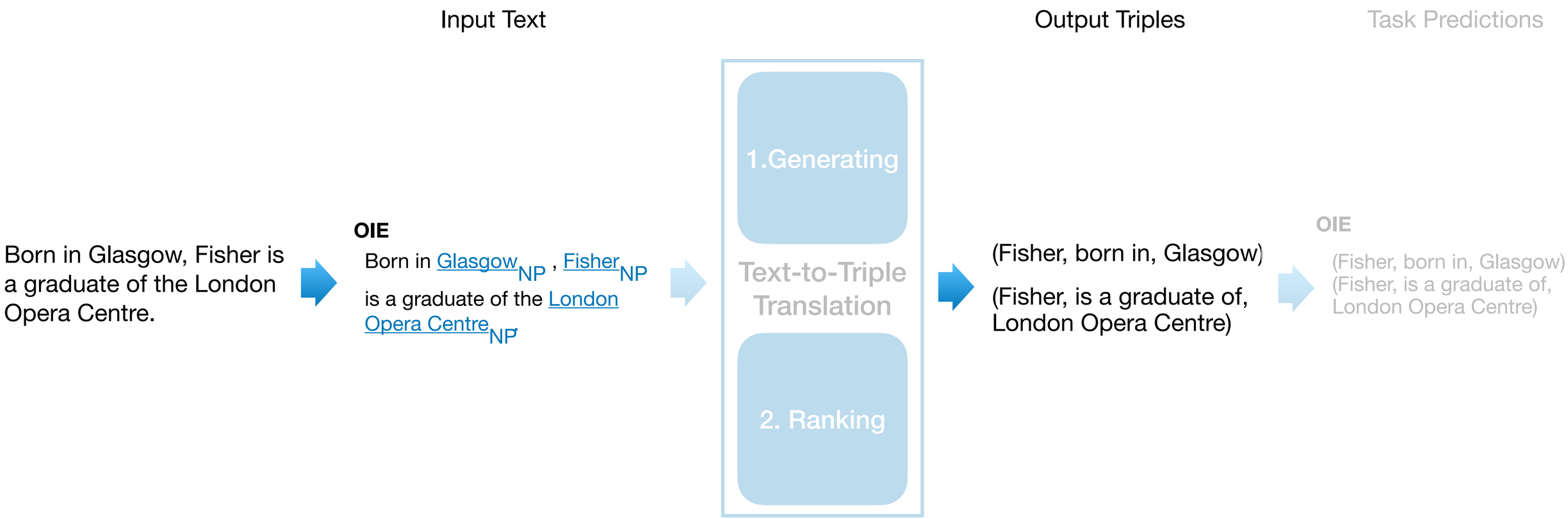


Same text-to-triple translation is shared across tasks, the only difference is the input encoding

An open information extraction (OIE) example



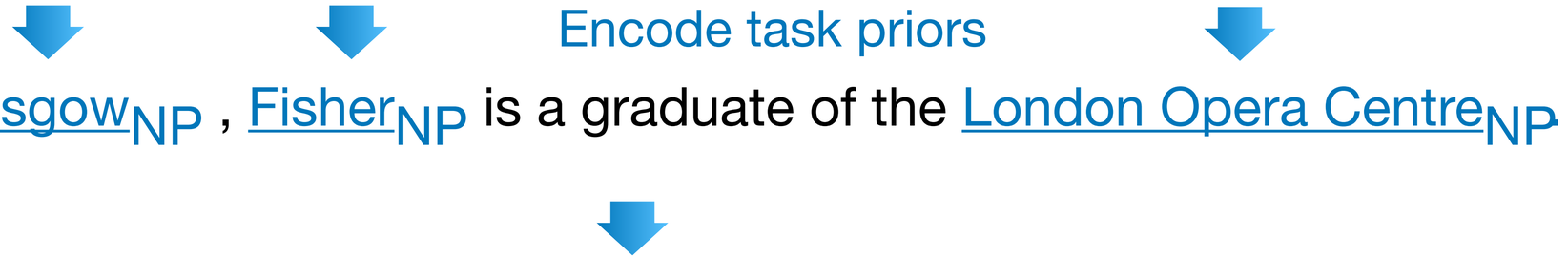
An OIE example: input and output format



An OIE example: input and output format

Input

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

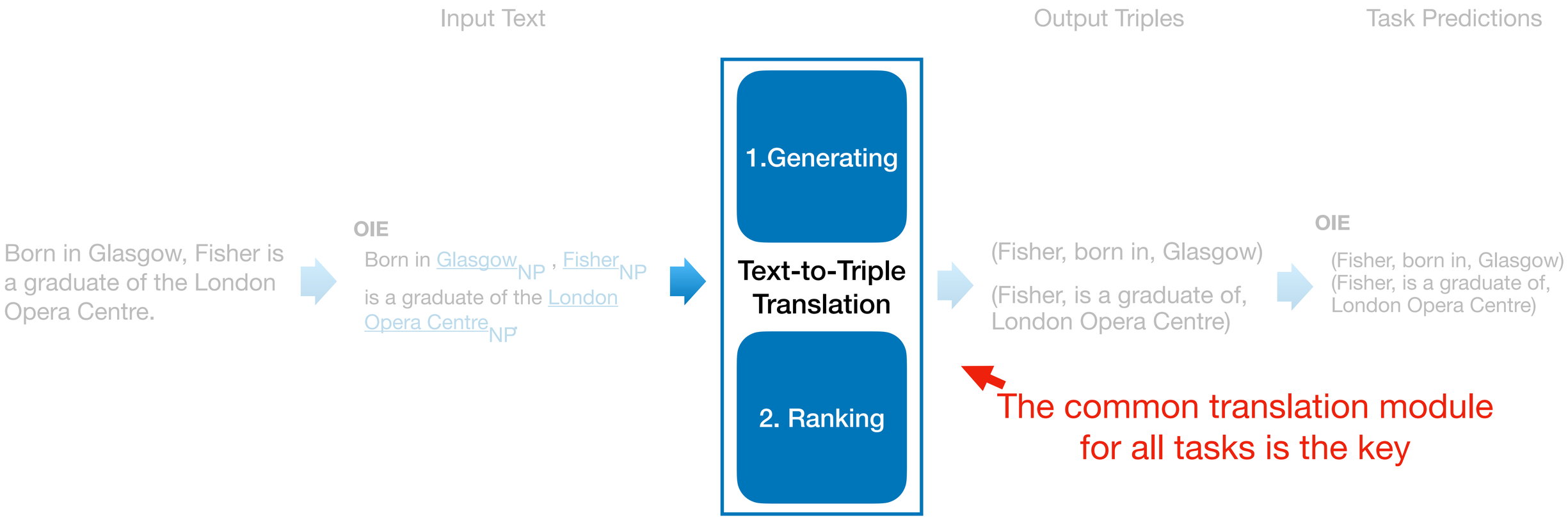


Output

(Fisher; Born in; Glasgow)
(Fisher; is a graduate of; London Opera Centre)

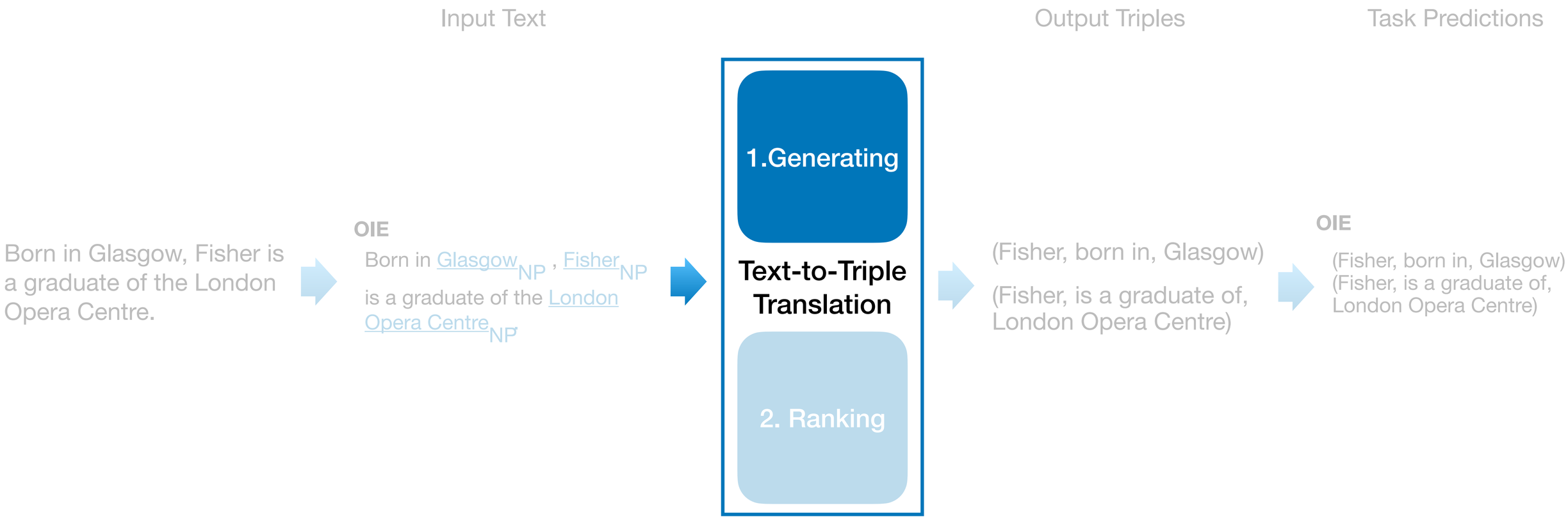
Input and output are designed in a format that is appropriate for OIE

An OIE example: zero-shot translation between input text and output triples



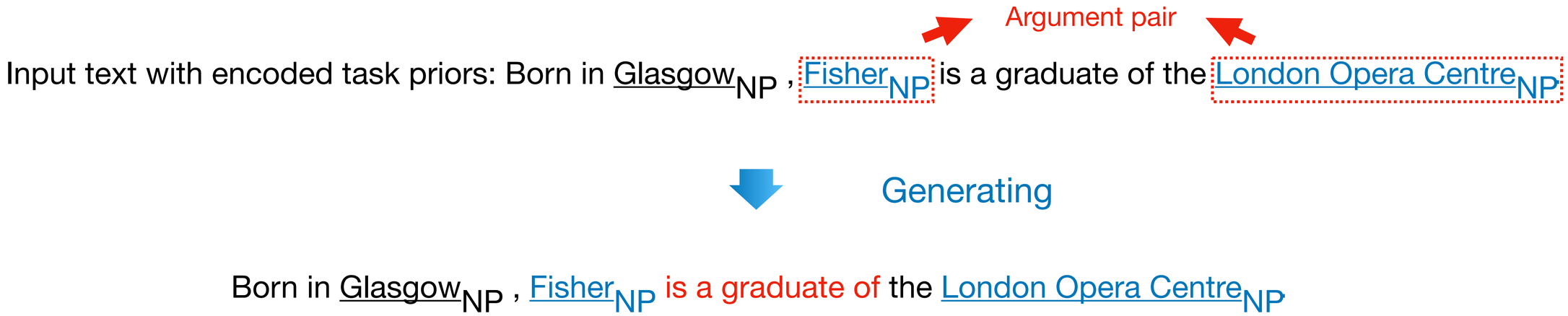
By leveraging the task priors encoded in the input, we enable the zero-shot transfer of the general knowledge that a pre-trained language model has about the task

An OIE example: generating triples from input text



An OIE example: generating triples from input text

OIE Formulation: Extract a set of sequences from input that are relevant to an argument pair

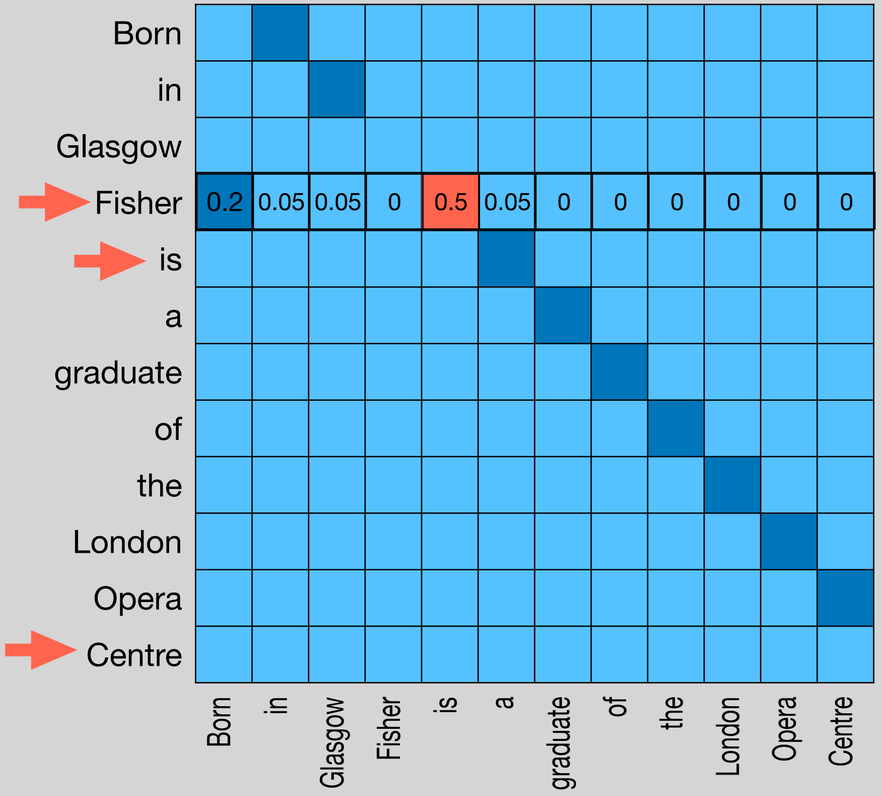


The generating stage produces general information about the task via pre-trained language models

An OIE example: generating triples from input text

Beam search with language model attention weights, beam size=1

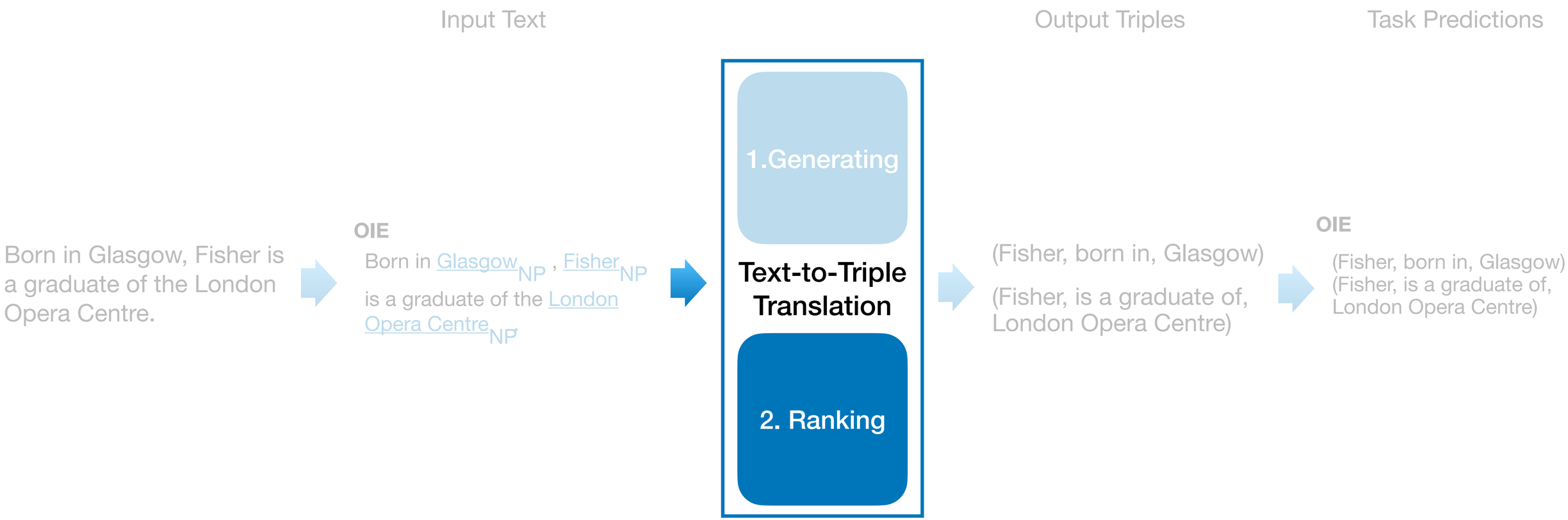
Input text with encoded task priors: Born in Glasgow$_{NP}$, Fisher$_{NP}$ is a graduate of the London Opera Centre$_{NP}$



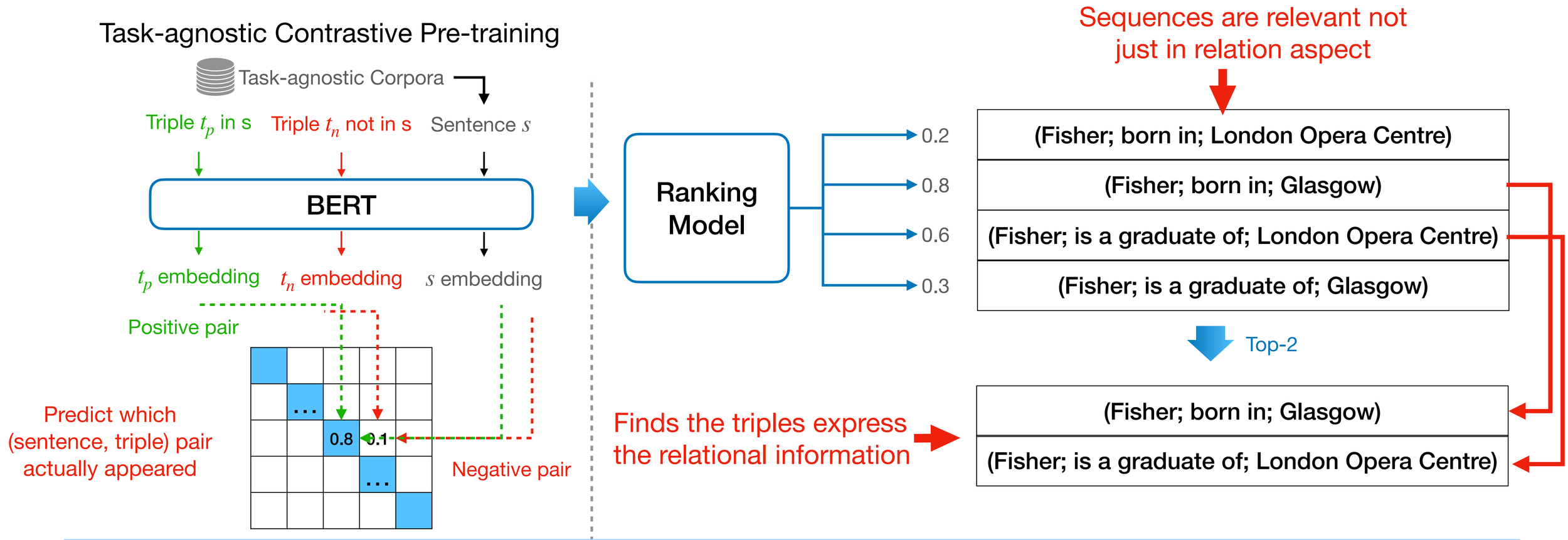
Step	Beam results
1	(Fisher;
2	(Fisher; is
3	(Fisher; is a
4	(Fisher; is a graduate
5	(Fisher; is a graduate of
.....	.....
9	(Fisher; is a graduate of; London Opera Centre)

Use the attention scores in pre-trained language models to measure the relevance between the sequence and the argument pair

An OIE example: ranking the generated triples

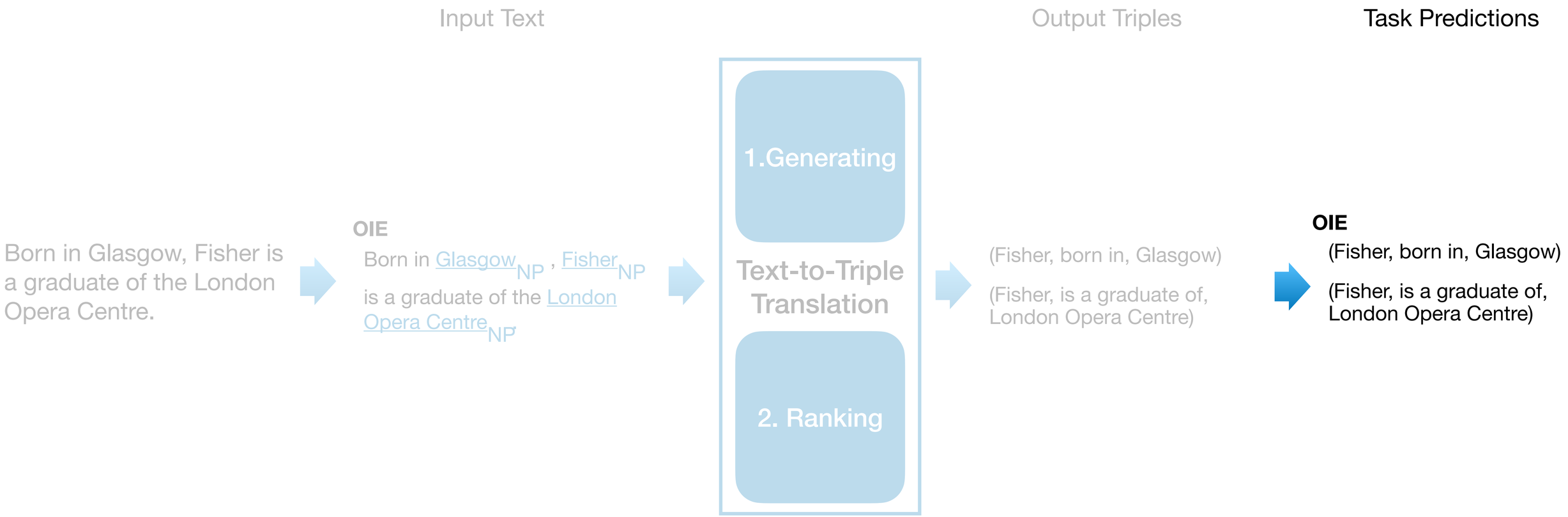


An OIE example: ranking the generated triples



The ranking stage finds triples that are of interest to the task via a ranking model pre-trained on a task-agnostic relational corpus

An OIE example: decoding task predictions from output triples



An OIE example: decoding task predictions from output triples

Output Triples

(Fisher; Born in; Glasgow)
(Fisher; is a graduate of; London Opera Centre)

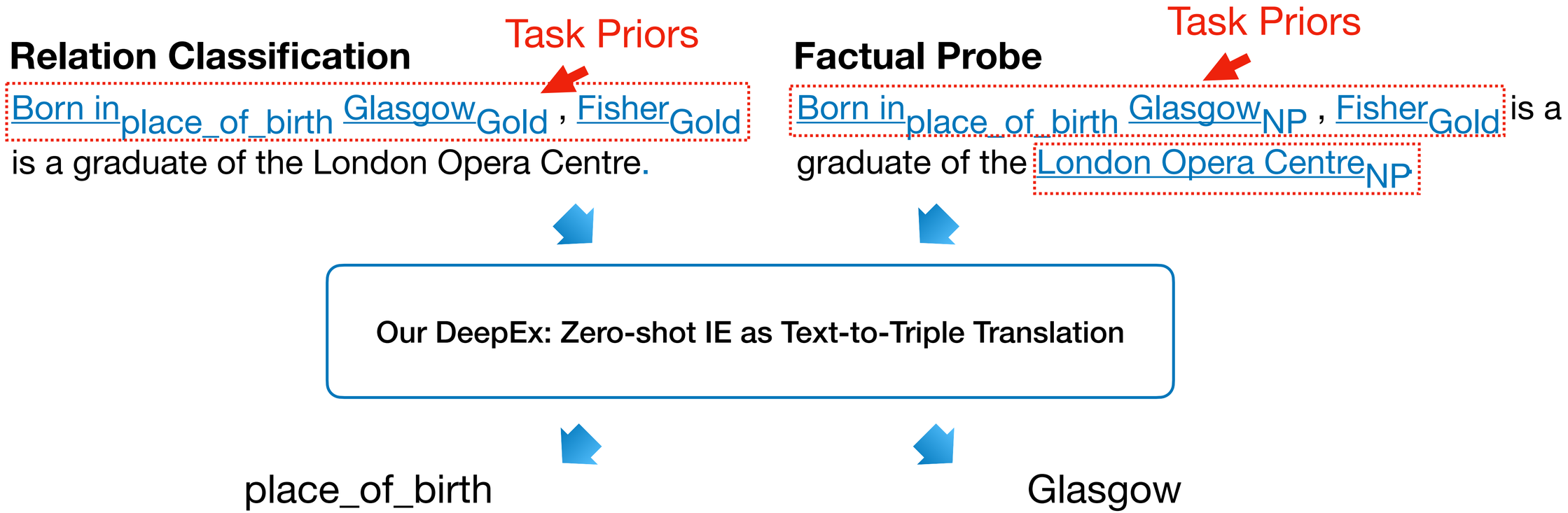
Decoding

Task Predictions

(Fisher; Born in; Glasgow)
(Fisher; is a graduate of; London Opera Centre)

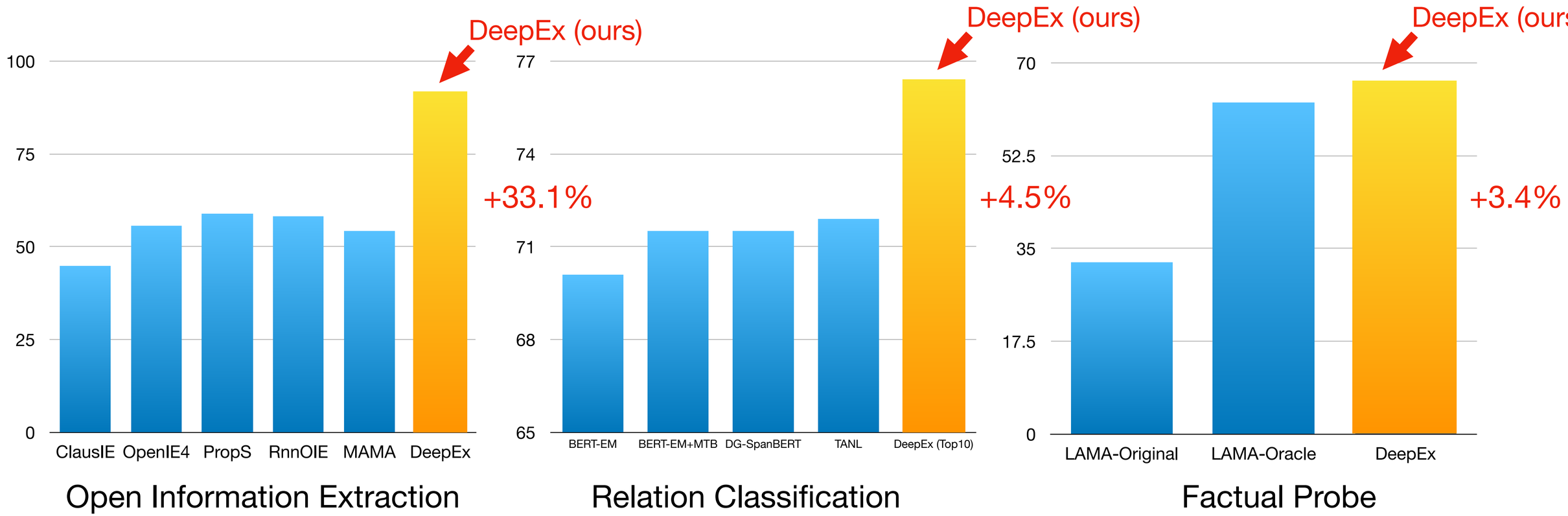
The framework encodes task priors in the input text and decodes the output triples to finally produce task predictions.

All information extraction tasks in the same framework



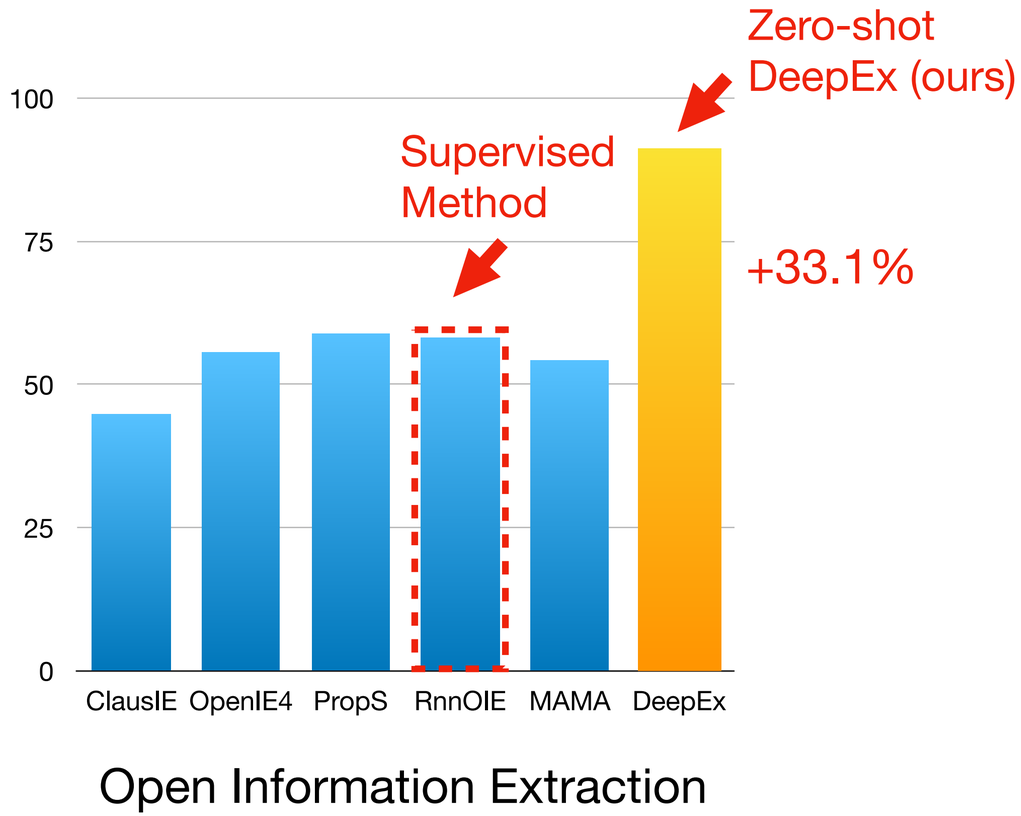
The framework encodes task priors in the input text and decodes the output triples to finally produce task predictions

Results: all three information extraction tasks

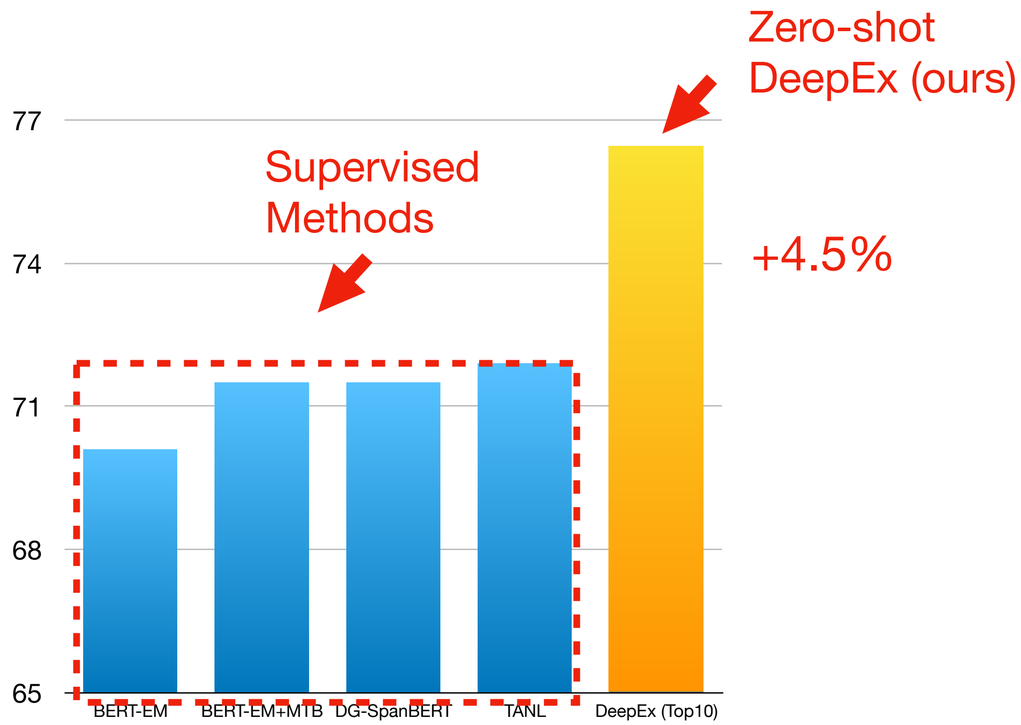


Our unified approach achieves state-of-the-art or competitive results on all tasks

Results: comparison between zero-shot (ours) and supervised performance



Open Information Extraction

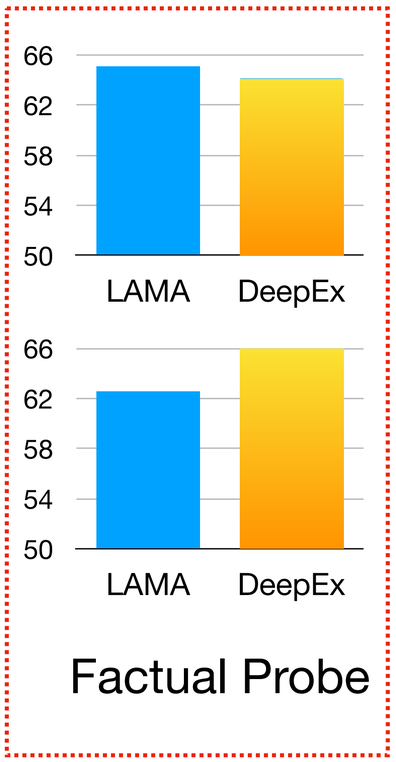


Relation Classification

Our zero-shot approach outperforms fully supervised task-specific models on open information extraction and relation classification

Results: comparison between interpretable (ours) and blackbox results

Comparable Results



Missing relations in sentences

Sentence: Judges' lodgings, the house once occupied by former Prime Minister Edward Heath at Salisbury.

Gold Triple: (Edward Heath; place_of_death; Salisbury)

Blackbox LAMA → **Salisbury** ✅

Interpretable DeepEx → ***[No prediction]*** ❌

Wrong memories of language models

Sentence: Nick Lucas's version, released on Brunswick, was a No.

Gold Triple: (Edward Heath; record_label; Brunswick)

Blackbox LAMA → **EMI** ❌

Interpretable DeepEx → ***Brunswick*** ✅

Our approach delivers more interpretable results due to enhanced model transparency

Conclusion

Unified framework that solves information extraction tasks

Competitive and state-of-the-art performance compared to fully supervised methods

DeepEx

Zero-shot information extraction without the need of any task-specific training set

Better interpretability through enhanced model transparency

Generalization by transferring the latent knowledge that language models have

Thank you for your time and interest!

Code: https://github.com/cgraywang/deepex