Adapters for Enhanced Modeling of Multilingual Knowledge and Text

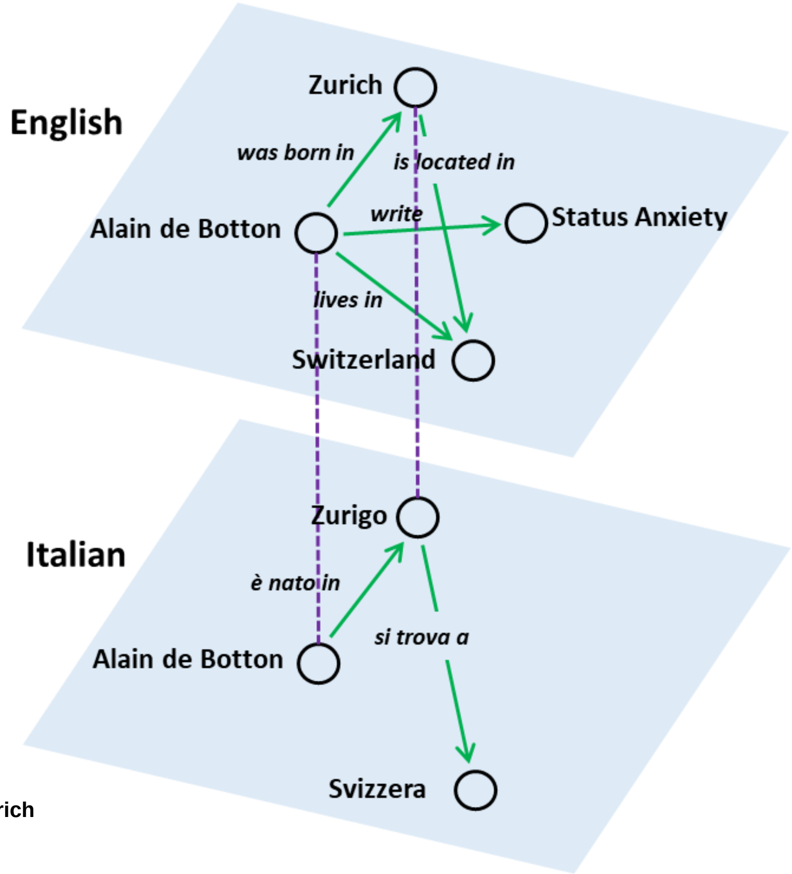
Yifan Hou, Wenxiang Jiao, Meizhen Liu, Carl Allen, Zhaopeng Tu, Mrinmaya Sachan

ETH zürich

Multilingual Knowledge Graph + Multilingual Language Model

(MLKG) (MLLM)

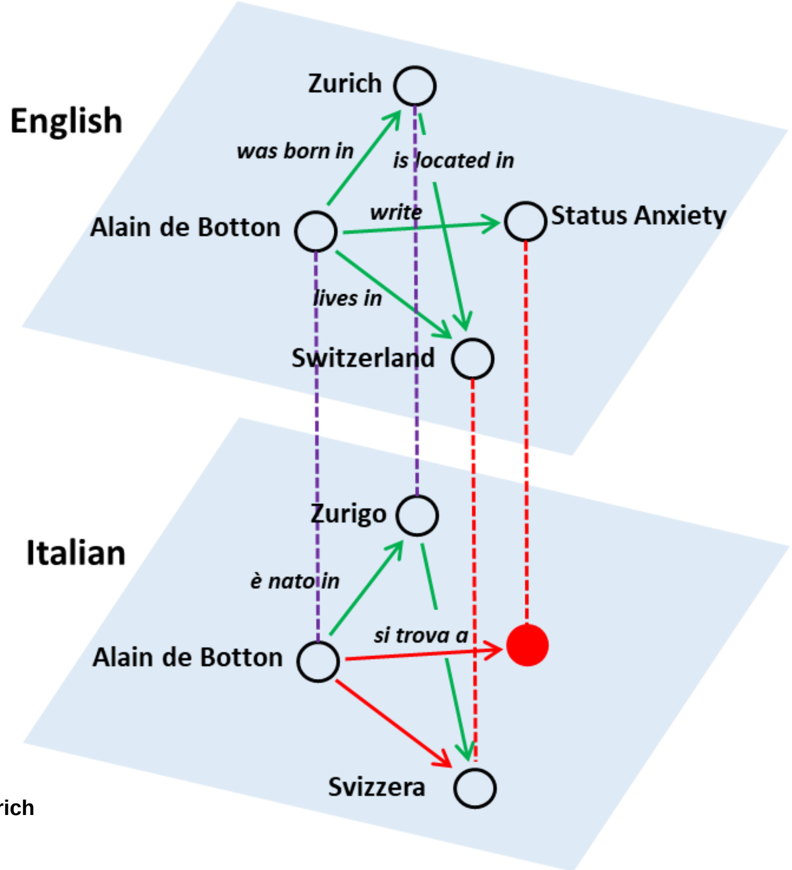
- MLKGs have **rich knowledge**



Multilingual Knowledge Graph + Multilingual Language Model

(MLKG) (MLLM)

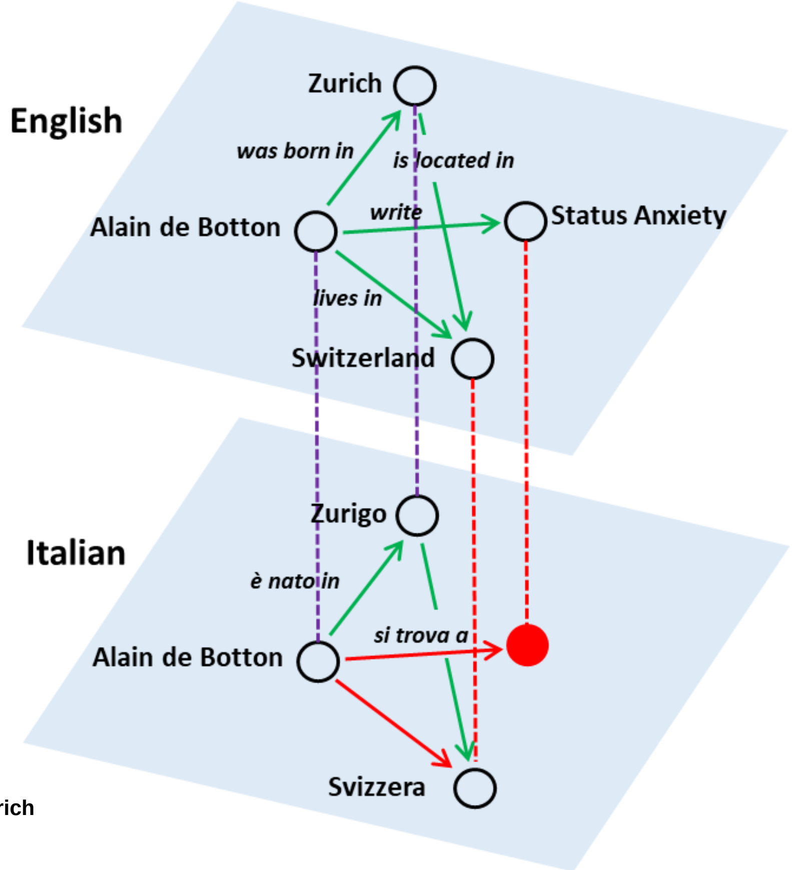
- MLKGs have **rich knowledge**
- But they are highly **incomplete**:
  - Missing triples / entities / alignments



Multilingual Knowledge Graph + Multilingual Language Model

(MLKG)

- MLKGs have **rich knowledge**
- But they are highly **incomplete**:
  - Missing triples / entities / alignments



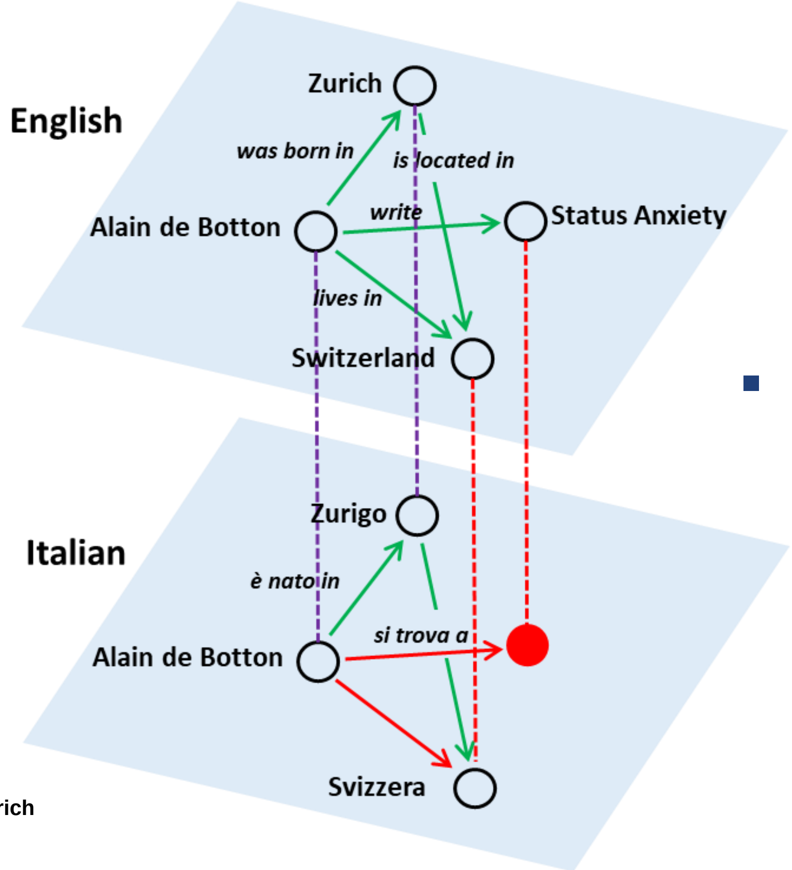
(MLLM)

- MLLMs have **strong transferability**
  - Transferring knowledge across 100+ languages
    - Devlin et al., 2019, Conneau et al., 2020
- But they **lack** factual/multilingual **knowledge**
  - Pretraining cannot capture much/many:
    - Sparse factual knowledge
    - Features of low resource languages

Multilingual Knowledge Graph + Multilingual Language Model

(MLKG)

- MLKGs have **rich knowledge**
- But they are highly **incomplete**:
  - Missing triples / entities / alignments



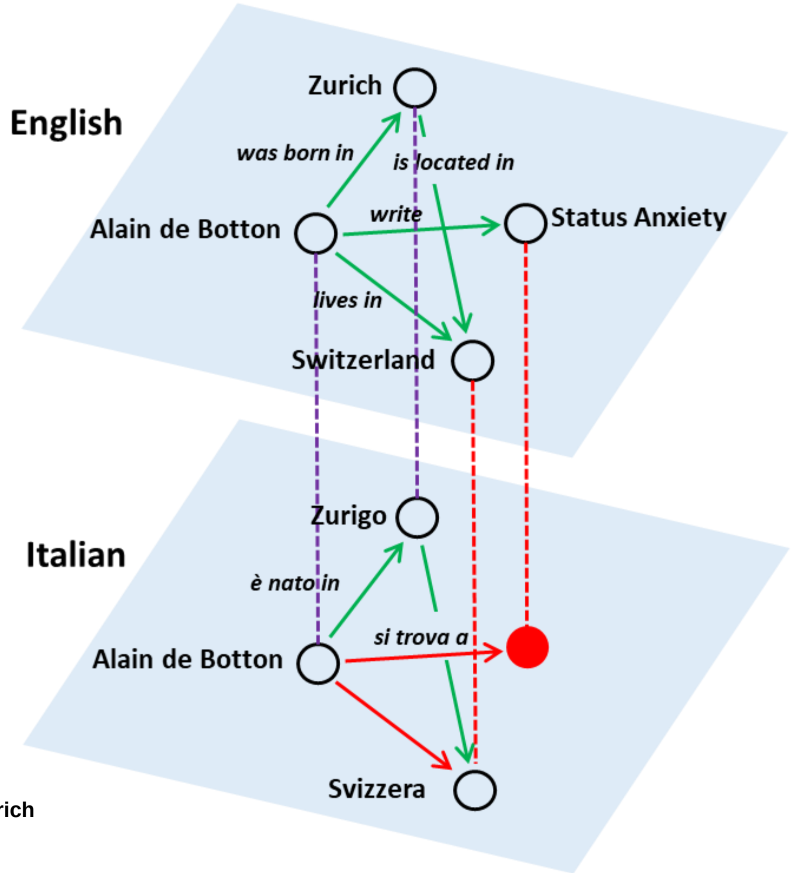
(MLLM)

- MLLMs have **strong transferability**
  - Transferring knowledge across 100+ languages
    - Devlin et al., 2019, Conneau et al., 2020
- But they **lack** factual/multilingual **knowledge**
  - Pretraining cannot capture much/many:
    - Sparse factual knowledge
    - Features of low resource languages

- Combining MLKG and MLLM ?
  - MLLM makes MLKG more **complete**
  - MLKG makes MLLM more "**knowledgeable**"

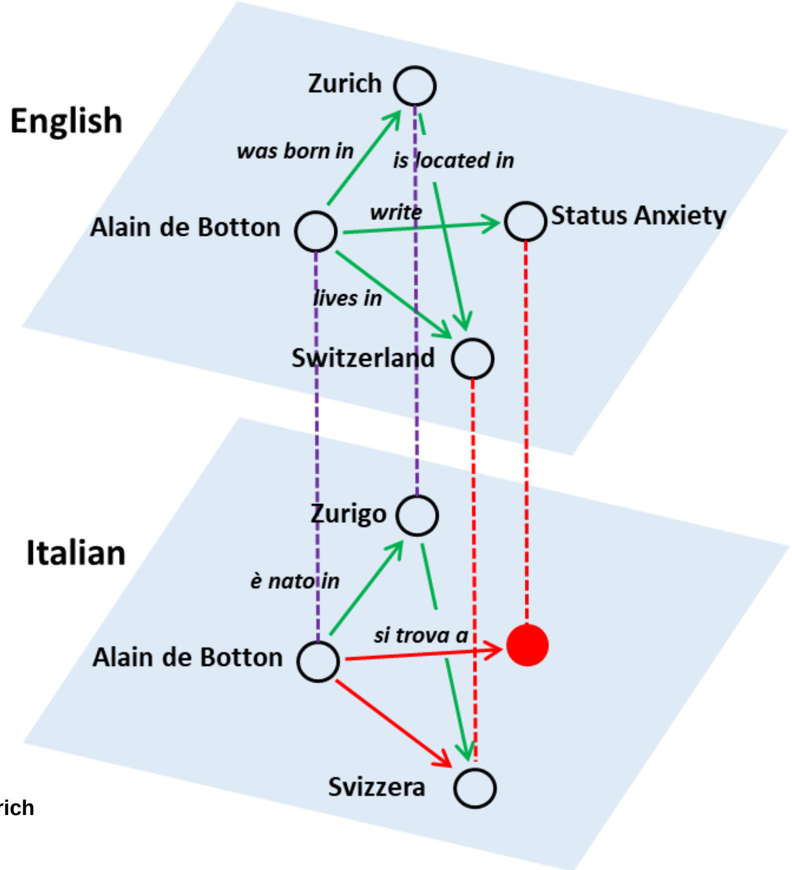
Knowledge Representations

- Multilingual knowledge graph (MLKG)



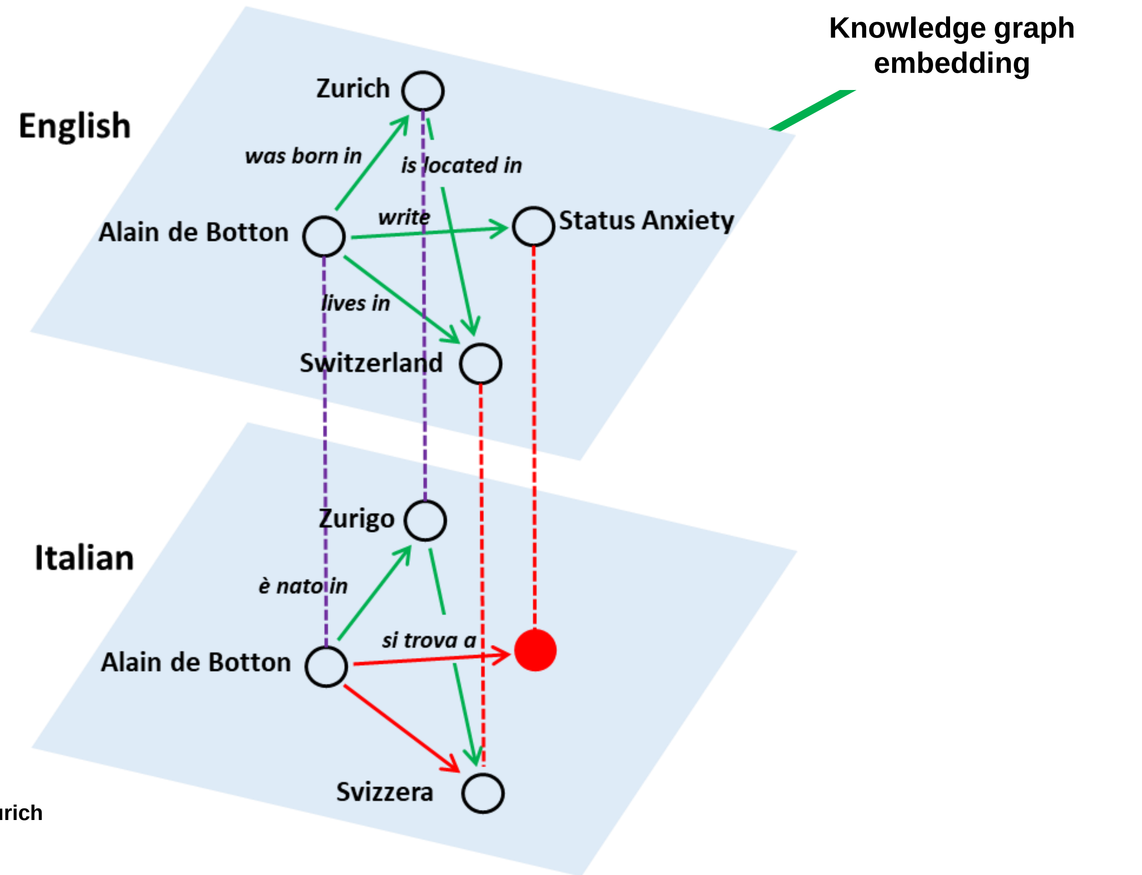
Knowledge Representations

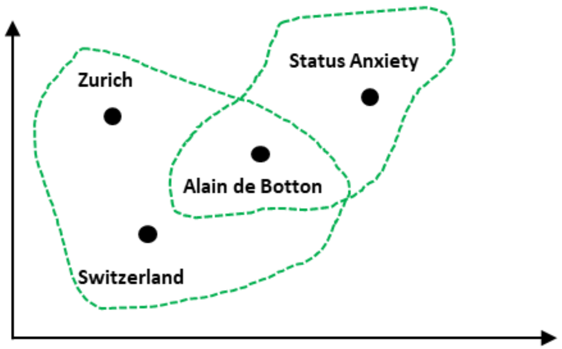
- Multilingual knowledge graph (MLKG)
  - Factual knowledge triples
  - Cross-lingual entity alignments



Knowledge Representations

- Multilingual knowledge graph (MLKG)
  - Factual knowledge triples
  - Cross-lingual entity alignments

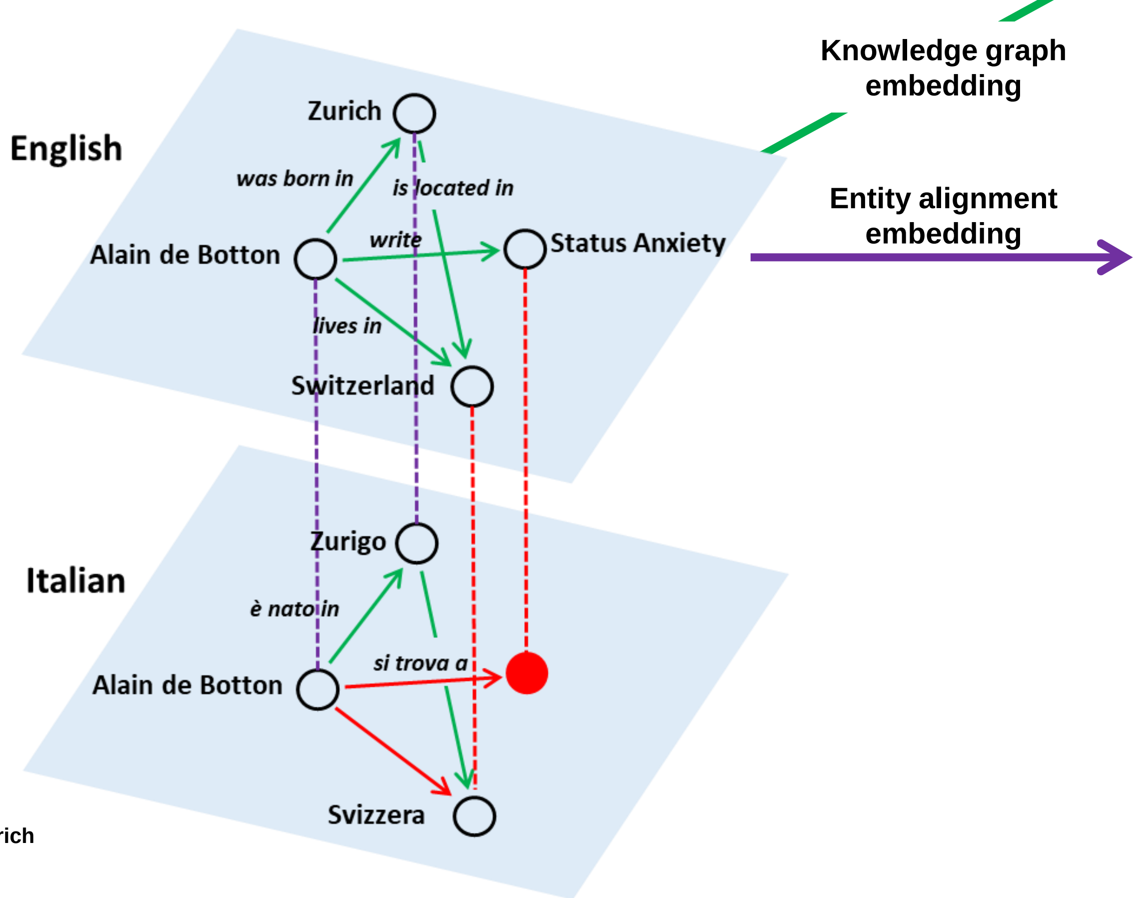


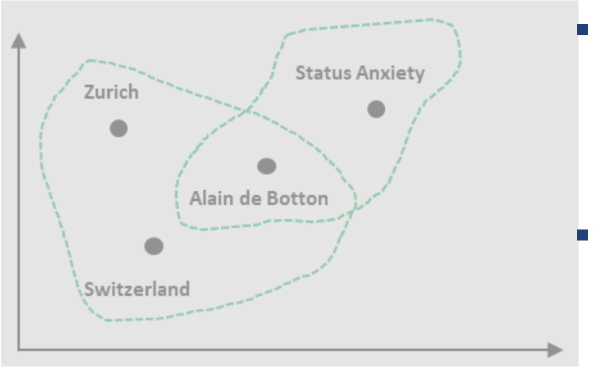


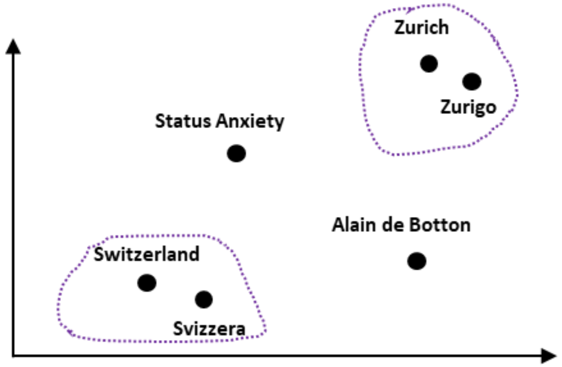
- Embedding objective:
  - h = head entity
  - r = relation
  - t = tail entity
- $||h + r - t||$
  - TransE (Bordes et al., 2013)

Knowledge Representations

- Multilingual knowledge graph (MLKG)
  - Factual knowledge triples
  - Cross-lingual entity alignments



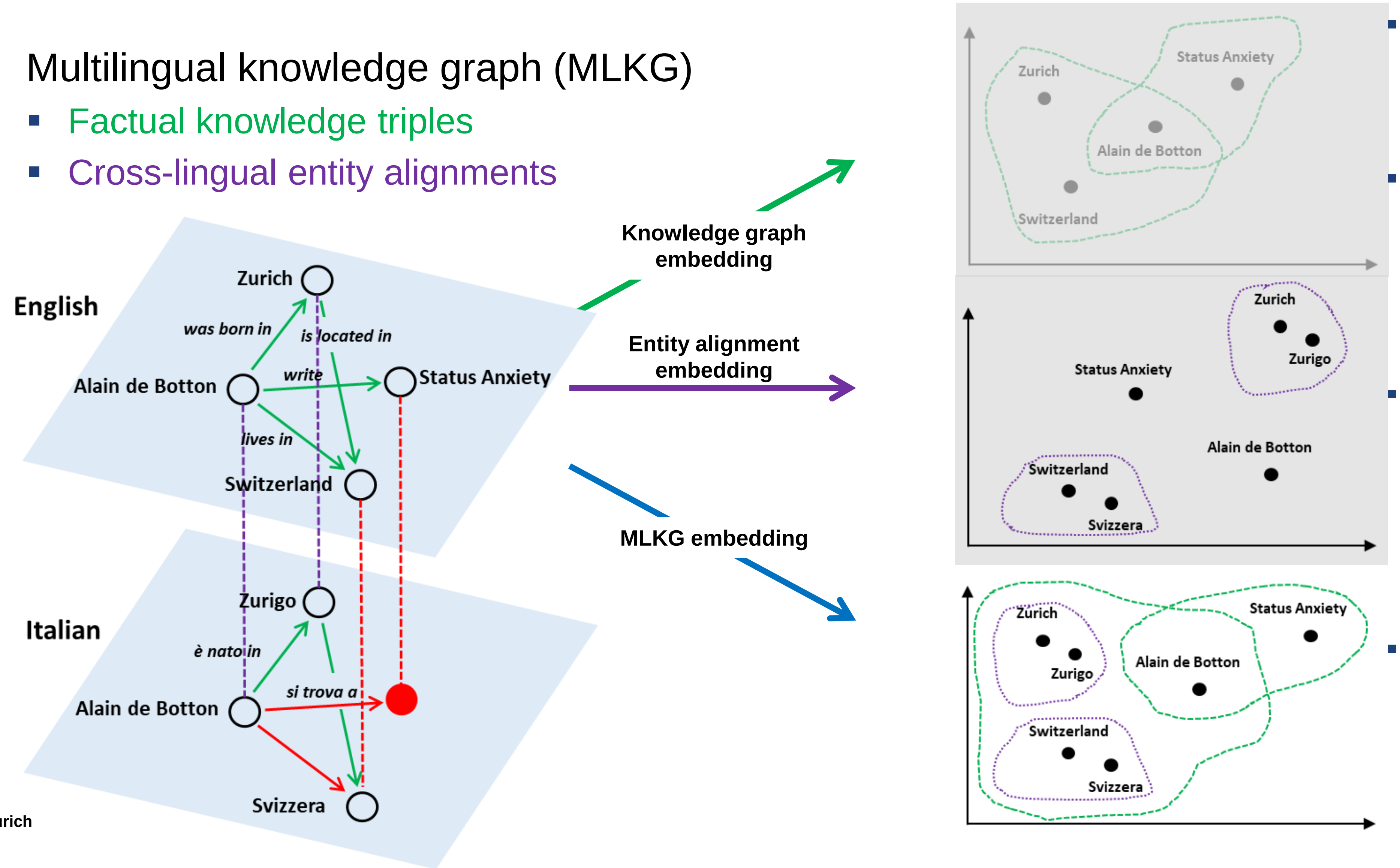




- Embedding objective:
  - h = head entity
  - r = relation
  - t = tail entity
- ||h + r - t||
  - TransE (Bordes et al., 2013)
- ||h - h’||

Knowledge Representations

- Multilingual knowledge graph (MLKG)
  - Factual knowledge triples
  - Cross-lingual entity alignments

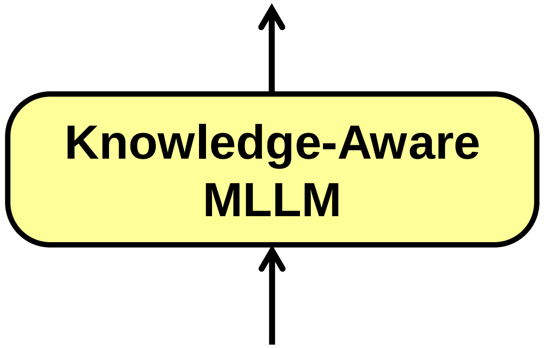


- Embedding objective:
  - h = head entity
  - r = relation
  - t = tail entity
- ||h + r - t||
  - TransE (Bordes et al., 2013)
- ||h - h'||
- ||h+r-t|| + ||h-h'|| + ||t-t'||
  - MTransE (Chen et al., 2017)

Knowledge-Aware Multilingual Language Model (MLLM)

- Knowledge in token representations
  - <u>Contextualized representation</u> $t_{\textbf{Switzerland}}$ should contain:

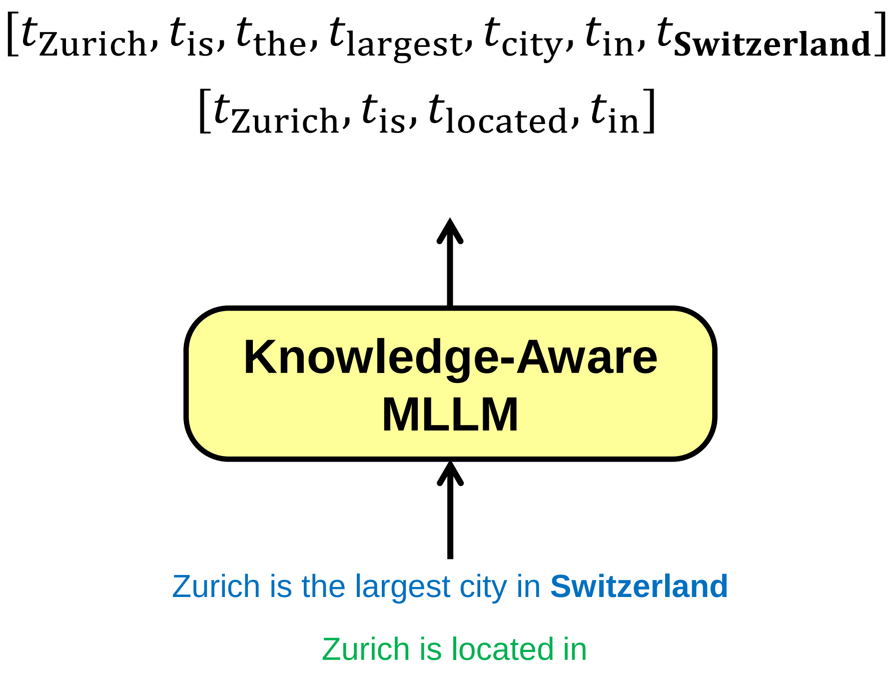
$$[t_{\text{Zurich}}, t_{\text{is}}, t_{\text{the}}, t_{\text{largest}}, t_{\text{city}}, t_{\text{in}}, t_{\textbf{Switzerland}}]$$

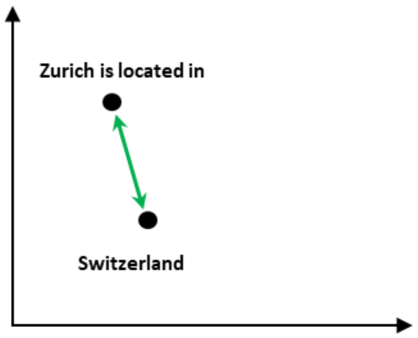


Zurich is the largest city in **Switzerland**

Knowledge-Aware Multilingual Language Model (MLLM)

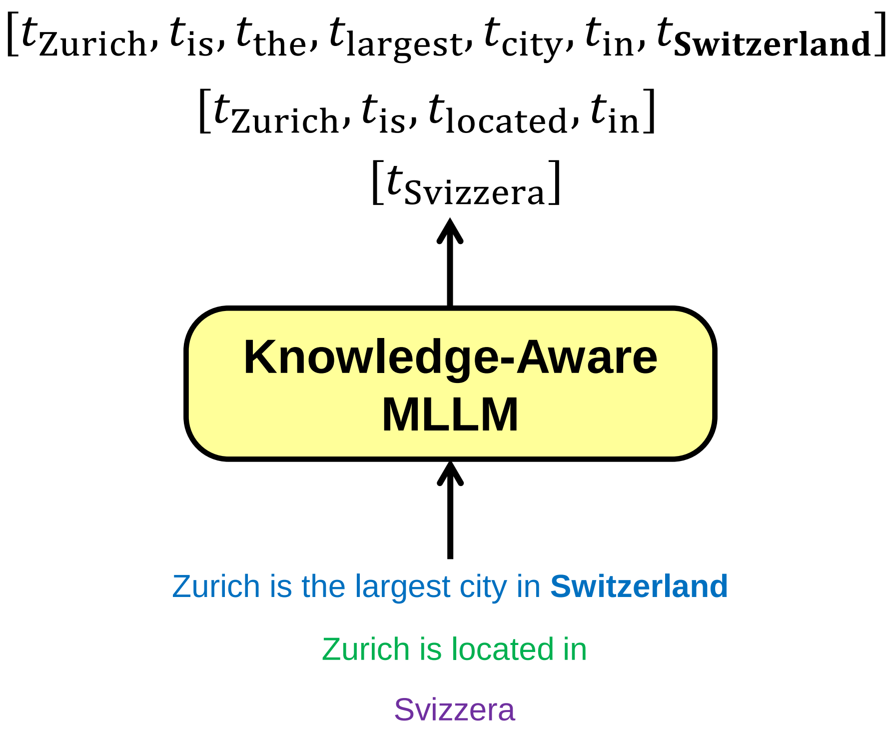
- Knowledge in token representations
  - Contextualized representation $t_{\mathbf{Switzerland}}$ should contain:
  - 1. Factual knowledge: (Zurich, is located in, **Switzerland**)
    - $\text{Average}(t_{\text{Zurich}}, t_{\text{is}}, t_{\text{located}}, t_{\text{in}}) \cong t_{\mathbf{Switzerland}}$
      - KnowBERT (Peters et al., 2019), ERNIE (Zhang et al., 2019)

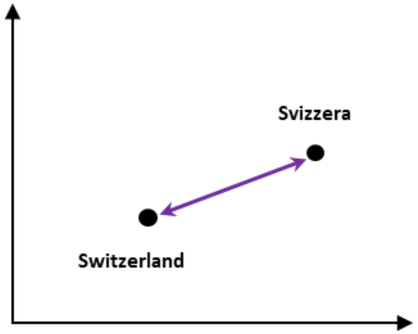




Knowledge-Aware Multilingual Language Model (MLLM)

- Knowledge in token representations
  - Contextualized representation $t_{\mathbf{Switzerland}}$ should contain:
  - 1. Factual knowledge: (Zurich, is located in, **Switzerland**)
    - $\text{Average}(t_{\text{Zurich}}, t_{\text{is}}, t_{\text{located}}, t_{\text{in}}) \cong t_{\mathbf{Switzerland}}$
      - KnowBERT (Peters et al., 2019), ERNIE (Zhang et al., 2019)
  - 2. Multilingual knowledge: (**Switzerland**, Svizzera)
    - $t_{\mathbf{Switzerland}} \cong t_{\text{Svizzera}}$
      - Universal semantic space





Knowledge Enhancement with Adapters

- Knowledgeable adapter set:
  - E/T => entity alignment / knowledge triple
  - P/S => phrase-level / sentence-level

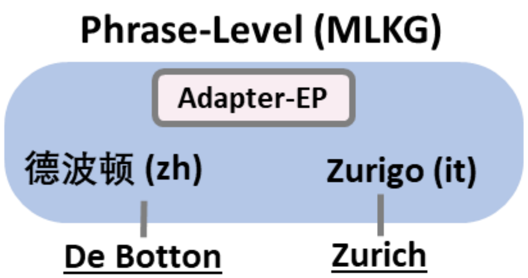
Task\Knowledge	Multilingual	Factual
MLKG	Adapter-EP	Adapter-TP
MLLM	Adapter-ES	Adapter-TS

Knowledge Enhancement with Adapters

- Knowledgeable adapter set:
  - E/T => entity alignment / knowledge triple
  - P/S => phrase-level / sentence-level

Task\Knowledge	Multilingual	Factual
MLKG	Adapter-EP	Adapter-TP
MLLM	Adapter-ES	Adapter-TS

  - Adapter-EP: MLKG entity alignment
    - Wikidata



Adapter Functions

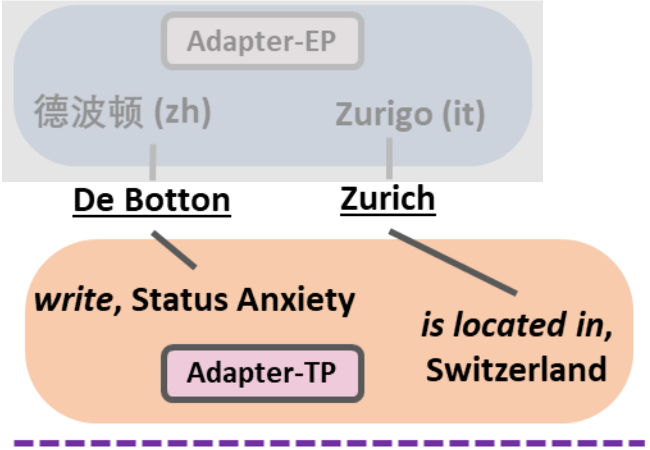
Knowledge Enhancement with Adapters

- Knowledgeable adapter set:
  - E/T => entity alignment / knowledge triple
  - P/S => phrase-level / sentence-level

Task\Knowledge	Multilingual	Factual
MLKG	Adapter-EP	Adapter-TP
MLLM	Adapter-ES	Adapter-TS

  - Adapter-EP: MLKG entity alignment
    - Wikidata
  - Adapter-TP: MLKG knowledge triples
    - Wikidata

Phrase-Level (MLKG)



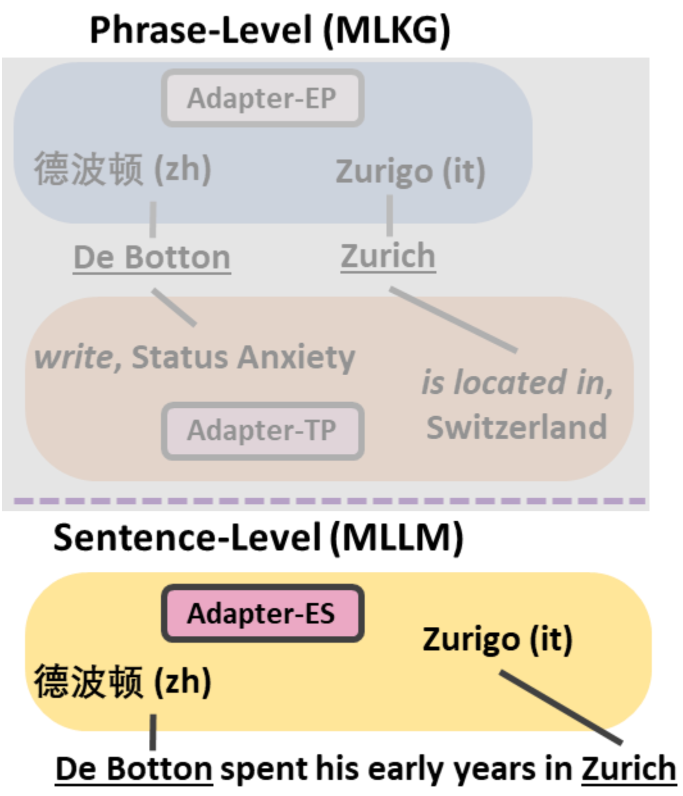
Adapter Functions

Knowledge Enhancement with Adapters

- Knowledgeable adapter set:
  - E/T => entity alignment / knowledge triple
  - P/S => phrase-level / sentence-level

Task\Knowledge	Multilingual	Factual
MLKG	Adapter-EP	Adapter-TP
MLLM	Adapter-ES	Adapter-TS

  - Adapter-EP: MLKG entity alignment
    - Wikidata
  - Adapter-TP: MLKG knowledge triples
    - Wikidata
  - Adapter-ES: Knowledge enhancement corpus
    - Wikipedia entity description



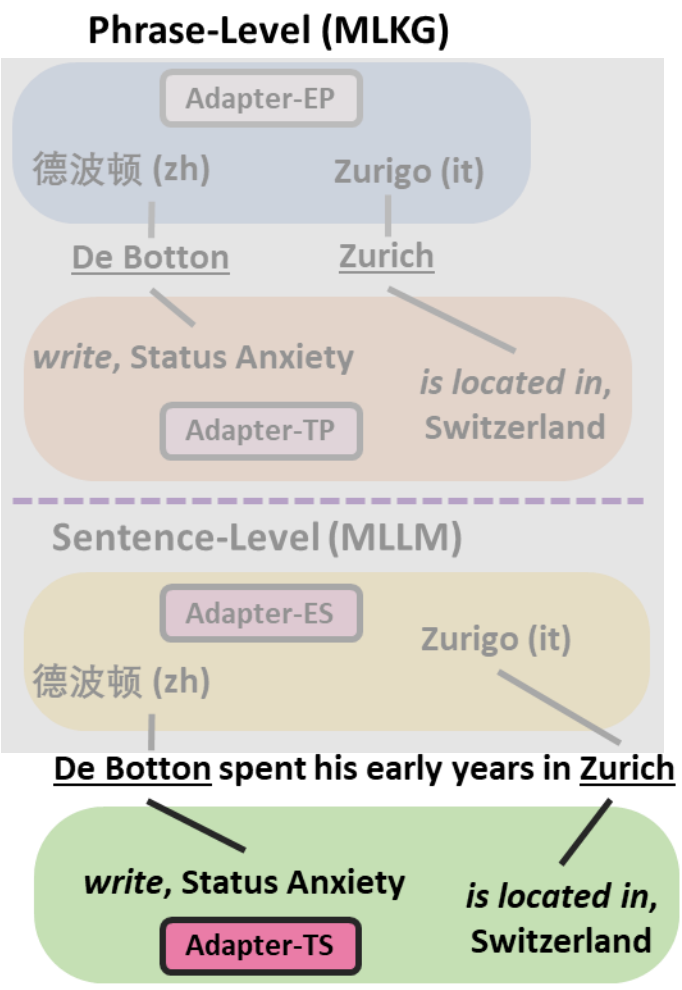
Adapter Functions

Knowledge Enhancement with Adapters

- Knowledgeable adapter set:
  - E/T => entity alignment / knowledge triple
  - P/S => phrase-level / sentence-level

Task\Knowledge	Multilingual	Factual
MLKG	Adapter-EP	Adapter-TP
MLLM	Adapter-ES	Adapter-TS

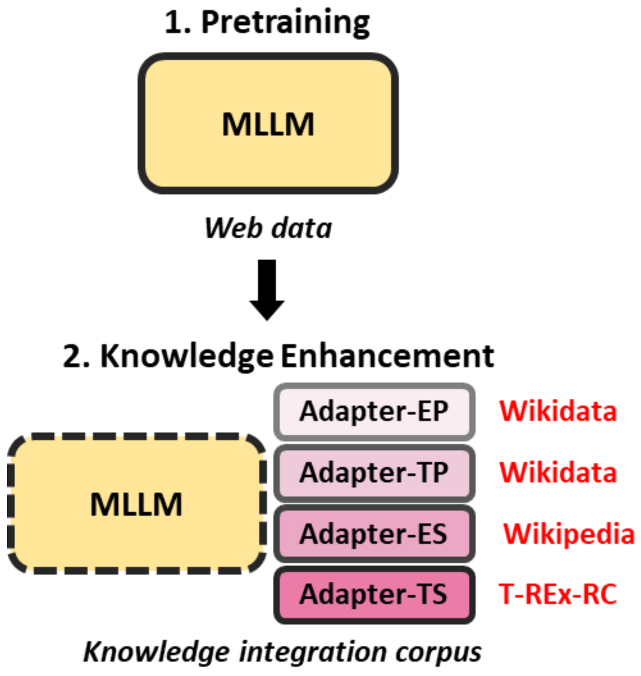
  - Adapter-EP: MLKG entity alignment
    - Wikidata
  - Adapter-TP: MLKG knowledge triples
    - Wikidata
  - Adapter-ES: Knowledge enhancement corpus
    - Wikipedia entity description
  - Adapter-TS: Knowledge enhancement corpus:
    - T-REx



Adapter Functions

Pipeline

- Adapter training (knowledge enhancement)
  - Training objectives: contrastive learning
    - InfoNCE loss (cosine) on MLLM output representations

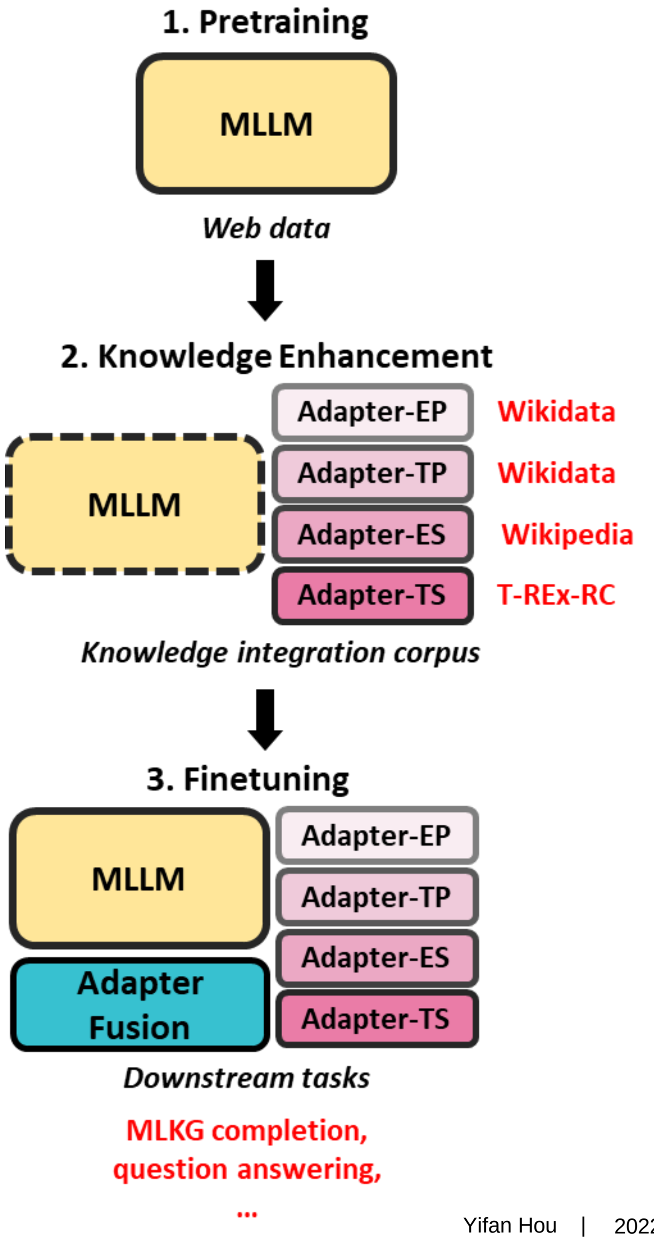


Pipeline

- Adapter training (knowledge enhancement)
  - Training objectives
    - InfoNCE loss (cosine) on MLLM output representations

- Finetuning whole enhanced MLLM on downstream tasks
  - MLLM, adapters, fusion module
  - Fusion Mechanism: attention aggregation
    - AdapterFusion (Pfeiffer et al., 2021)

- Inference

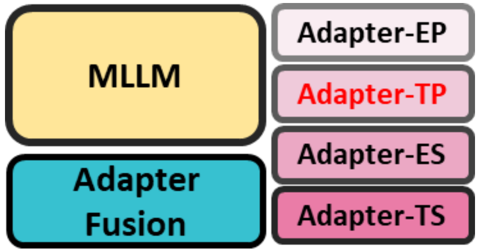


Results: MLKG Completion

- **Knowledge triple completion**
  - Given head entity label and relation in **one language**, find the tail entity
    - E.g., Italian **(Zurigo, si trova in, Svizzera)**
      - (Zurich, is located in, Switzerland)

Input head entity and relation (it):
Zurigo, si trova in

All candidate tail entity (it):
[Svizzera, De Botton, …]

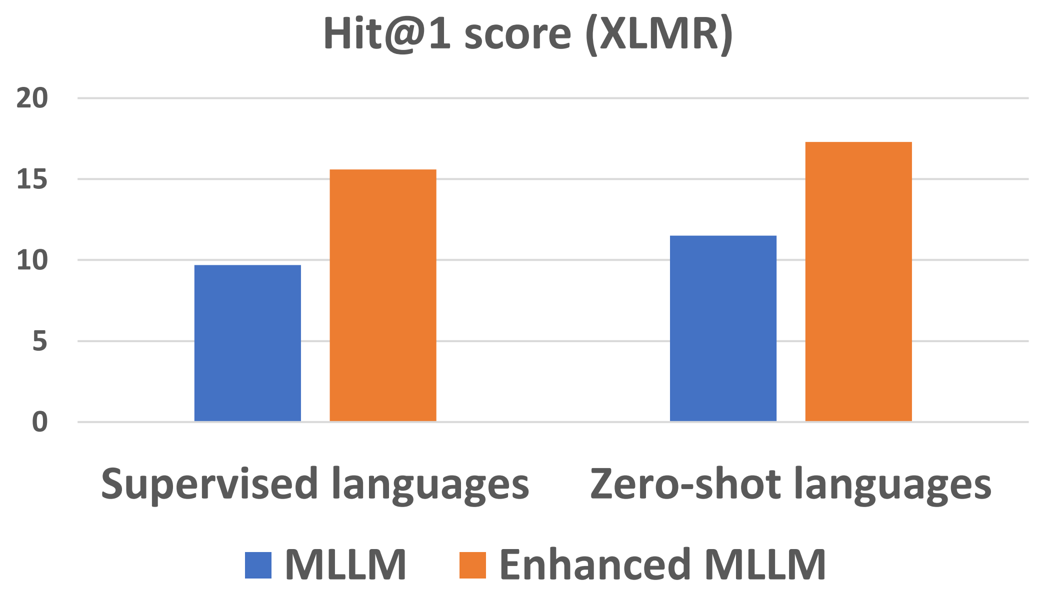


Match

(Zurigo, si trova in, Svizzera)

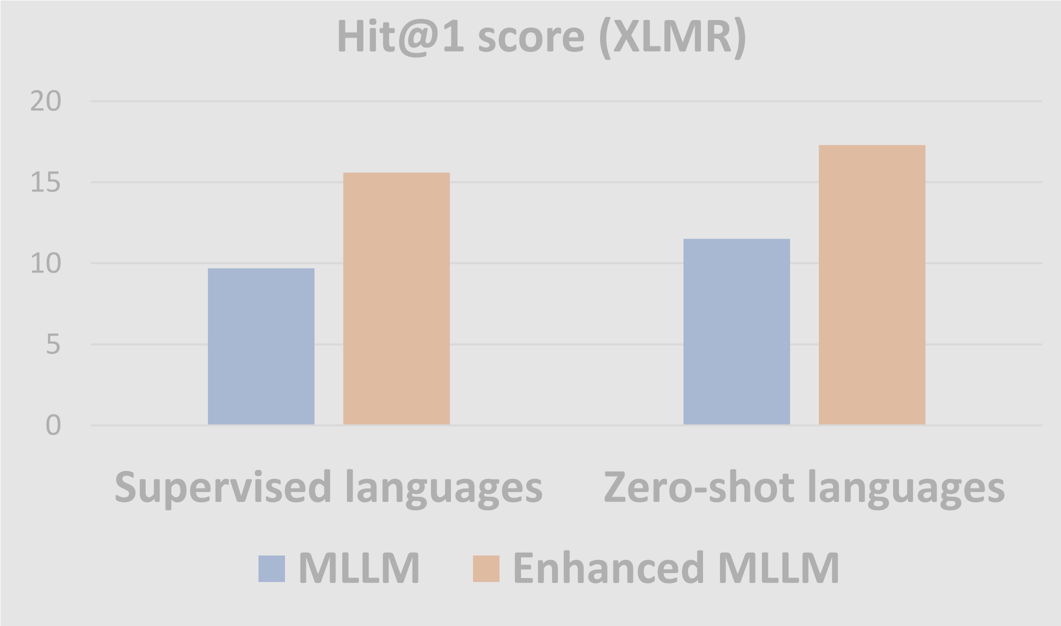
Results: MLKG Completion

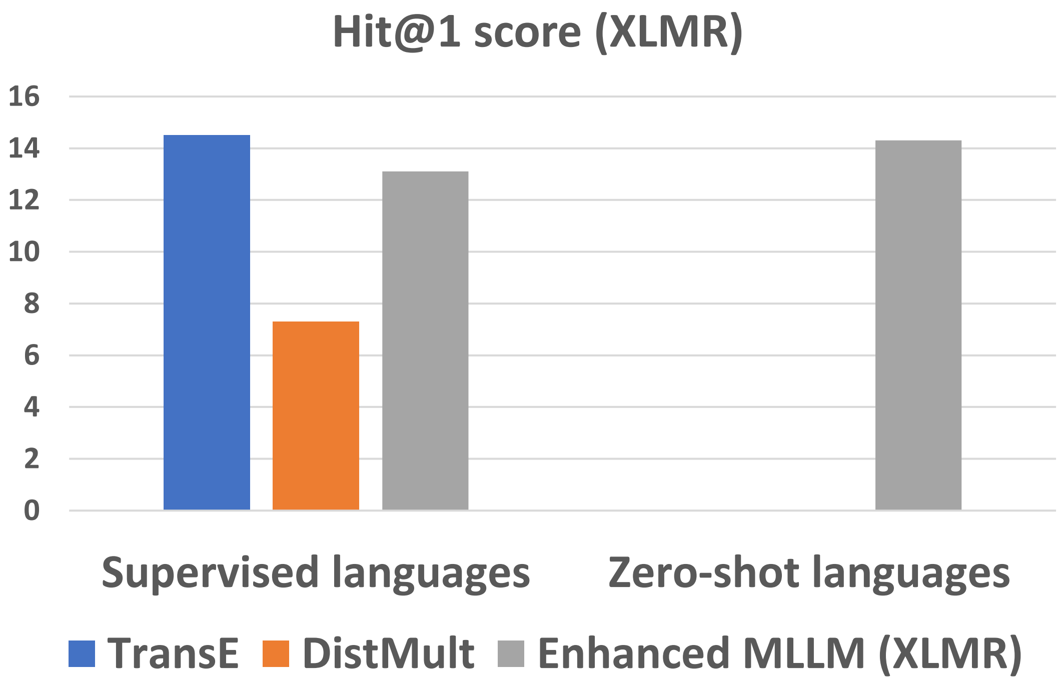
- **Knowledge triple completion**
  - 1. Enhanced MLLMs always outperform base MLLMs



Results: MLKG Completion

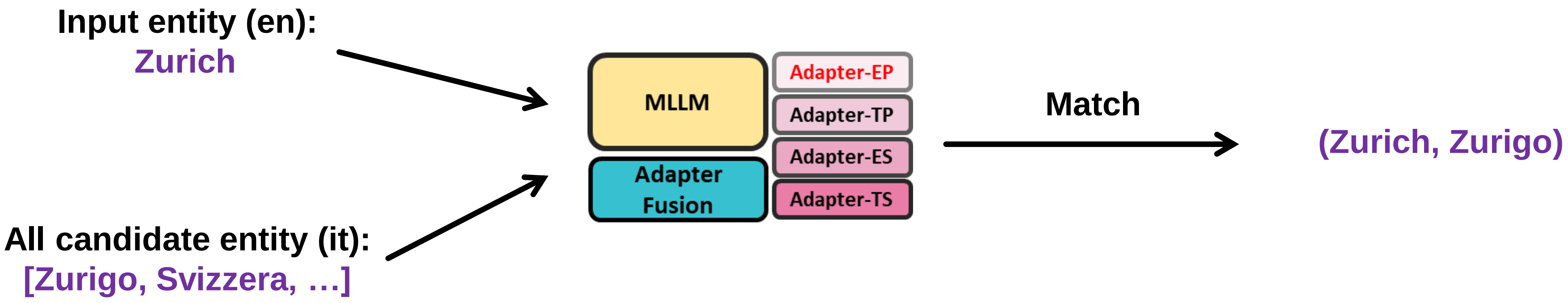
- **Knowledge triple completion**
  - 1. Enhanced MLLMs always outperform base MLLMs
  - 2. Comparable to existing baselines
    - Especially for zero-shot languages
    - Existing baselines cannot support





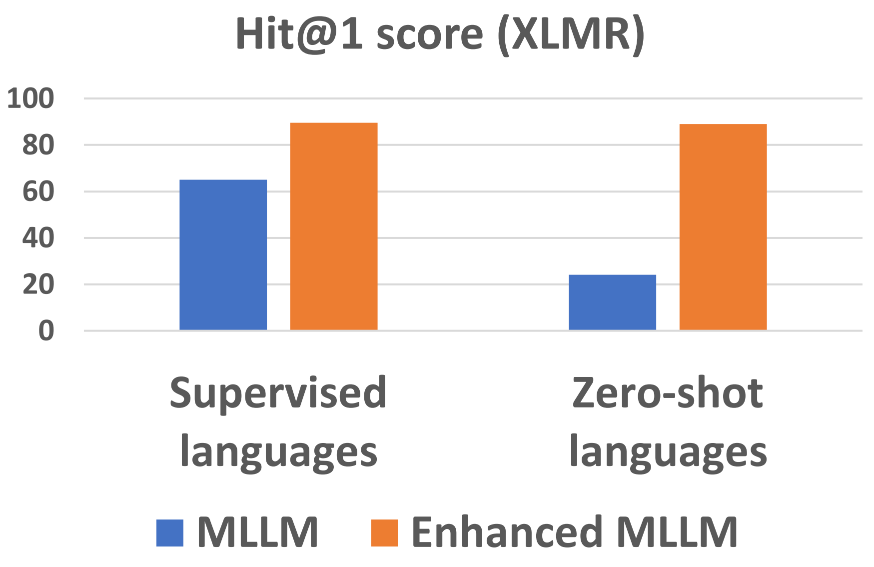
Results: MLKG Completion

- **Cross-lingual entity alignment**
  - Given entity label in **English**, find aligned one in **other language**
    - E.g., Italian **(Zurich, Zurigo)**



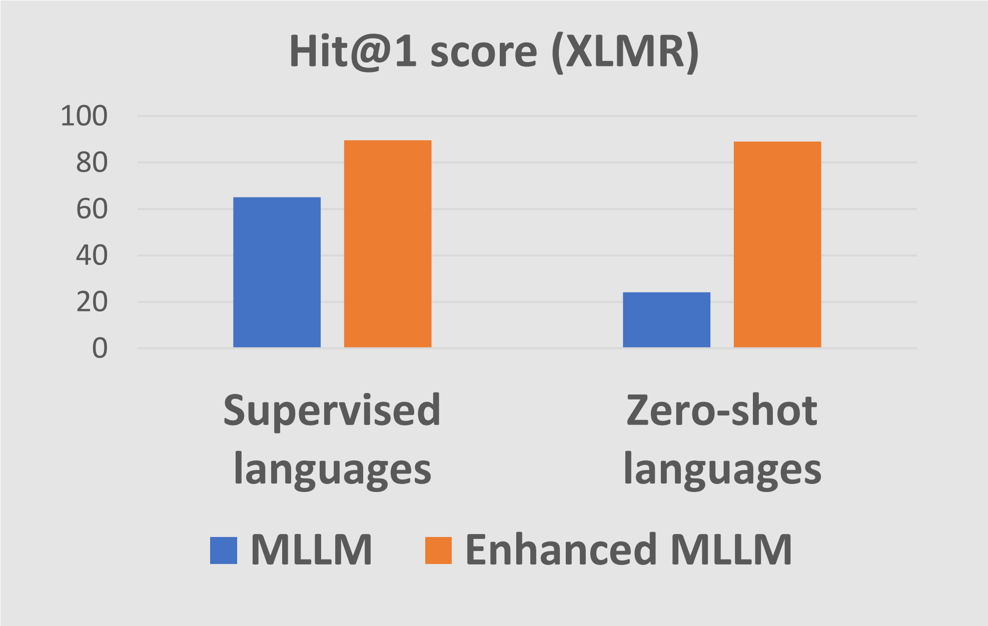
Results: MLKG Completion

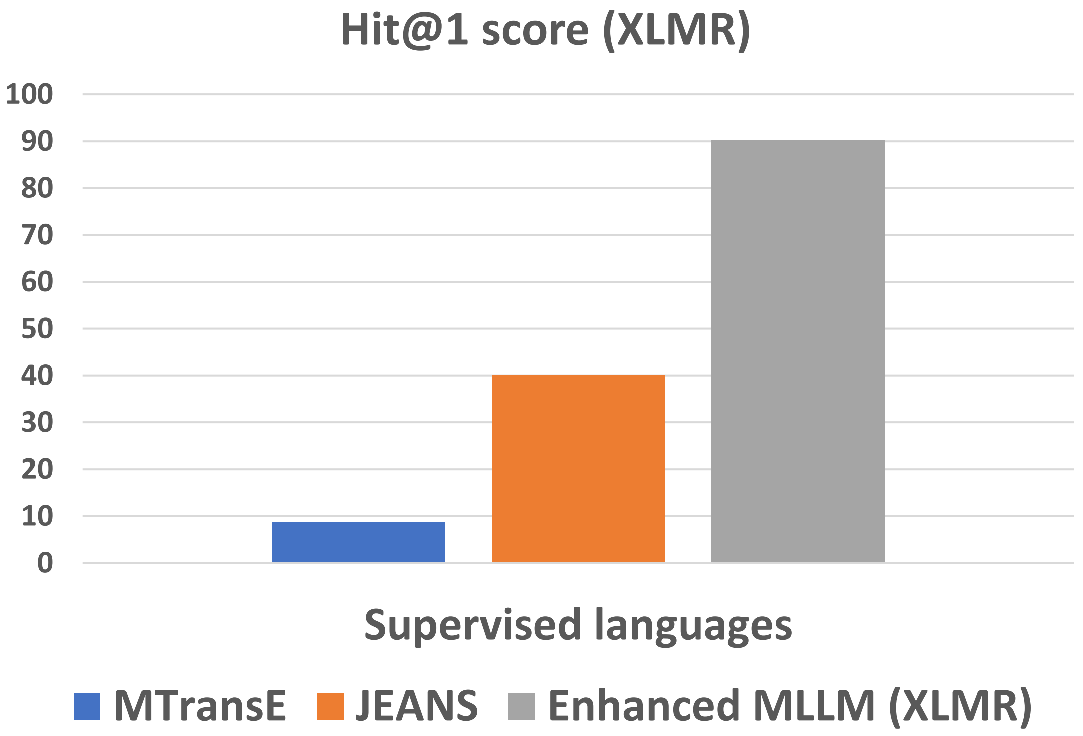
- **Cross-lingual entity alignment**
  - 1. Enhanced MLLMs always outperform base MLLMs
    - Especially for zero-shot languages



Results: MLKG Completion

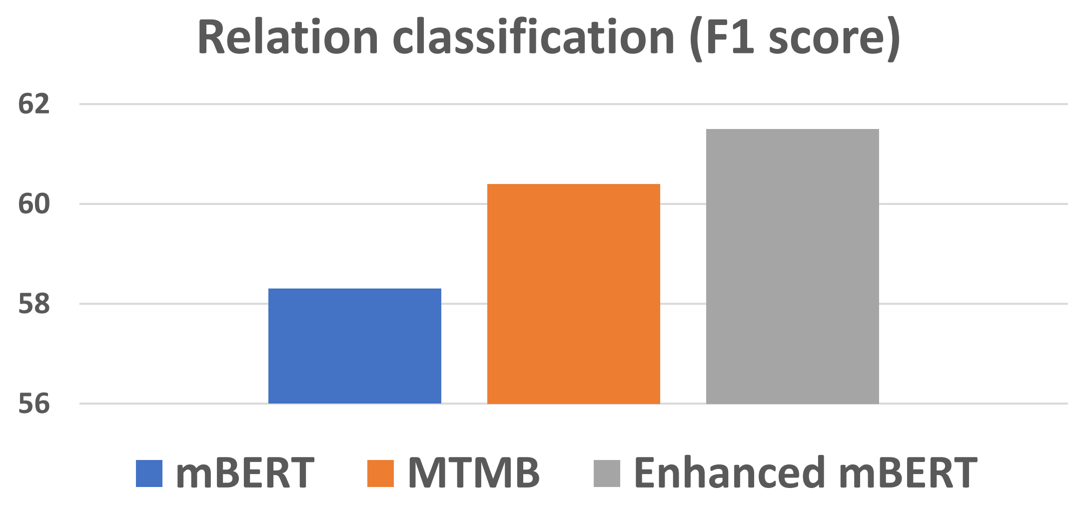
- **Cross-lingual entity alignment**
  - 1. Enhanced MLLMs always outperform base MLLMs
    - Especially for zero-shot languages
  - 2. Much better than previous baselines
    - E.g., JEANS





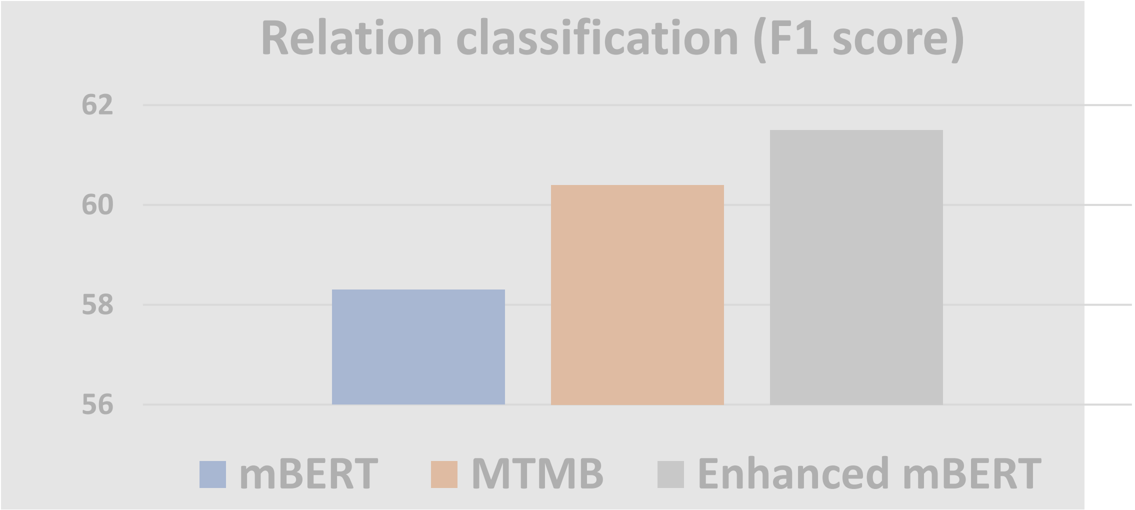
Results: MLLM Tasks

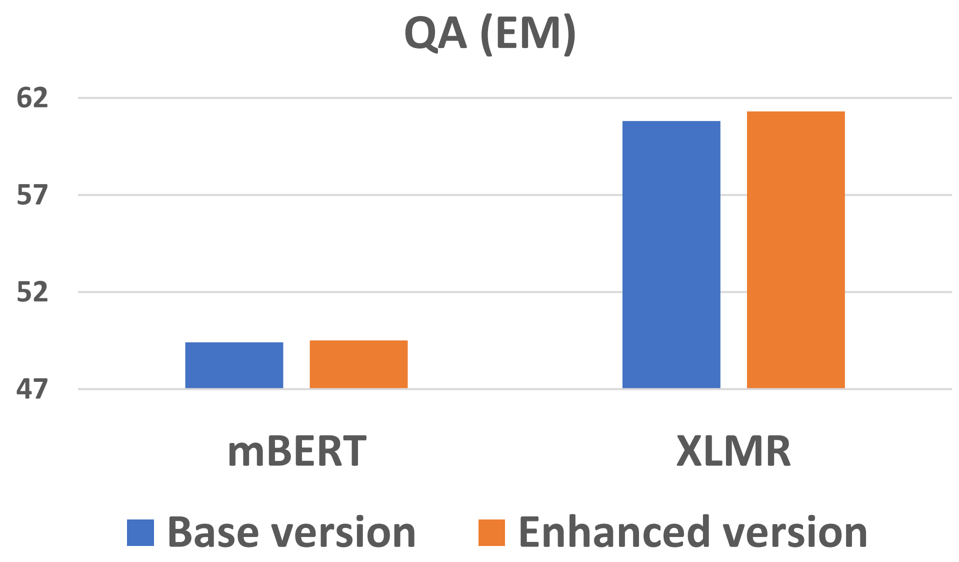
- **Knowledge enhancement**
  - Knowledge-intensive task
    - Relation classification (Köksal and Özgür, 2020)



Results: MLLM Tasks

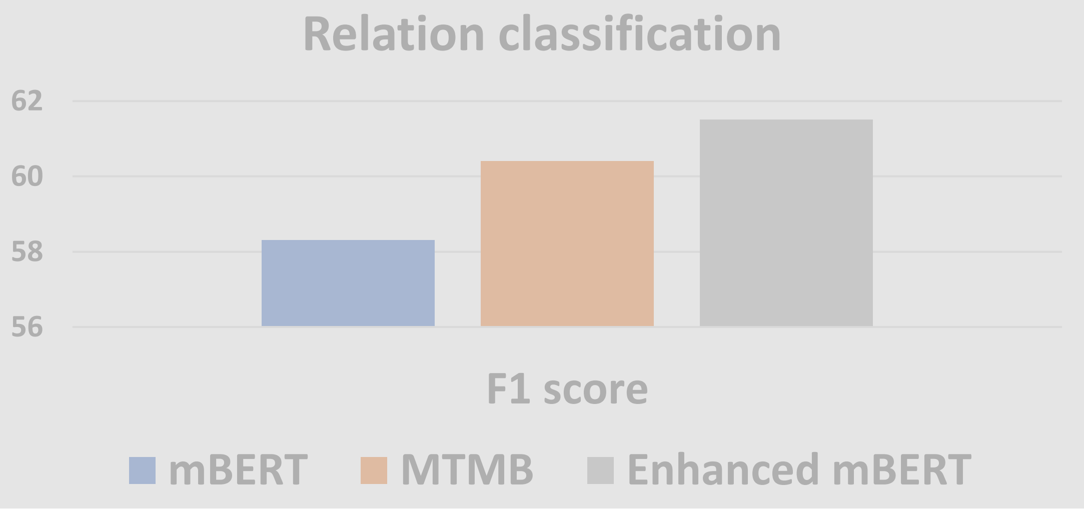
- **Knowledge enhancement**
  - Knowledge-intensive task
    - Relation classification (Köksal and Özgür, 2020)
  - General language modelling tasks
    - Question Answering (SQuAD & XQuAD)

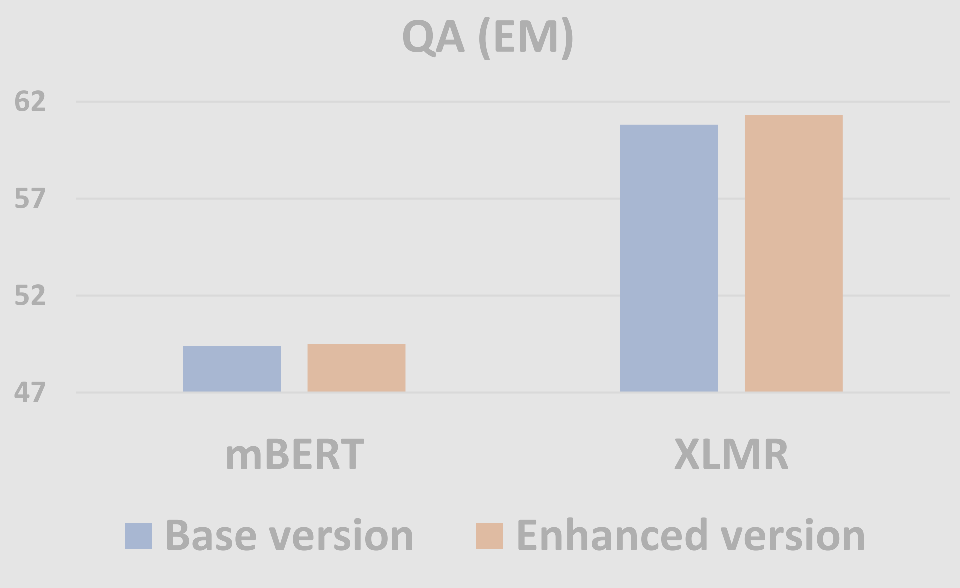


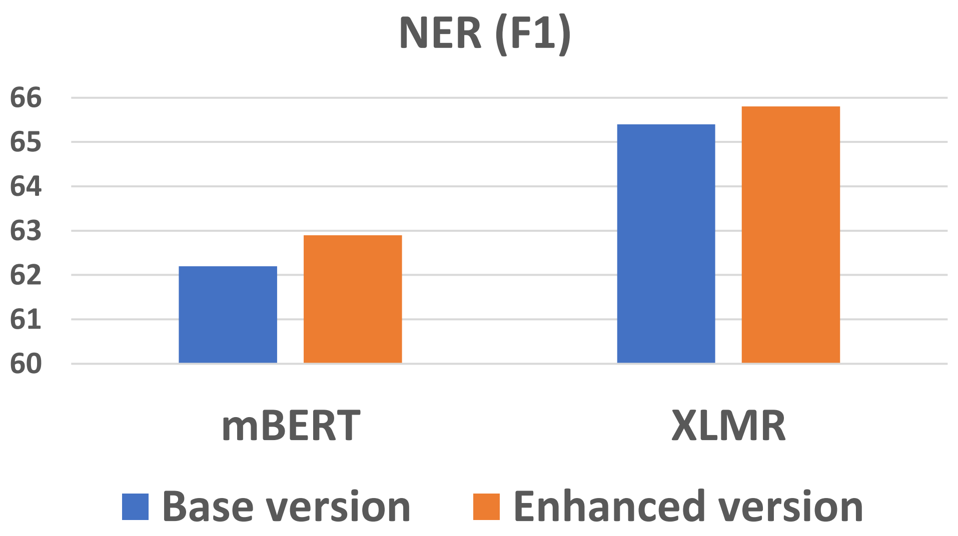


Results: MLLM Tasks

- **Knowledge enhancement**
  - Knowledge-intensive task
    - Relation classification (Köksal and Özgür, 2020)
  - General language modelling tasks
    - Question Answering (SQuAD & XQuAD)
    - Name Entity Recognition (WikiAnn)







Takeaways

- 1. Combining MLKG and MLLM benefit modeling of both multilingual knowledge and text
  - MLKGs become more complete
  - MLLMs become more knowledgeable
- 2. Enhancement with adapters and contrastive learning works good

All trained adapters are now available on AdapterHub

Code, datasets, and extended benchmarks

Thanks!