Semantically Consistent LLM-Based Text Generation (SemLLM)

Claire Gardent, CNRS/LORIA

IA Grand Est
ENACT

Conditional Generation: The generated text must be faithful to the Input

Patient Data

Age 45

Sex Male

Symptoms Persistent cough

Diagnosis Pneumonia

Treatment Antibiotics

INCORRECT

21 y.o. female with a headache due to a migraine is given antibiotics.

45 y.o. male with a cough due to pneumonia is given amoxicillin.

INCOMPLETE

45 y.o. male with a cough due to pneumonia

CORRECT

45 y.o. male with a cough due to pneumonia is given antibiotics.

Improving and Evaluating Faifthfulness, OOD Generalisation

Generation

Data → Text

- Multilingual Verbalisation of Structured Data.
- Improving Generalisation and Faithfulness
- Evaluating Faithfulness

Text → Data

- Cross-lingual Biography Generation.
- Reducing Data and Model Biases

Analysis

Text → Text

- Extracting temporal and spatial information form the Notre-Dame de Paris data science project.

People

Yifei Song

PhD Candidate

Started April 2025

EPFL Master

Kun Zhang

Postdoctoral Fellow

Started September 2025

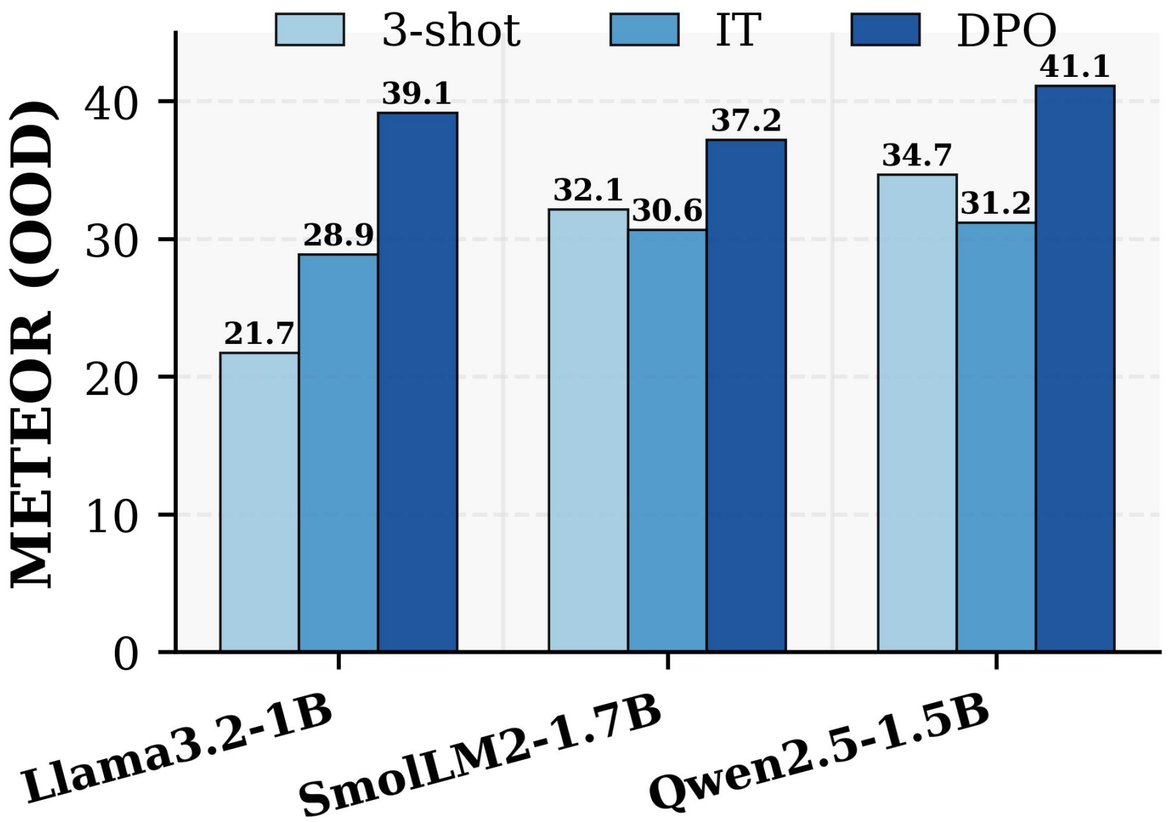
PhD INRIA Saclay/LIX

MuCAL: Contrastive Alignement for Preference Driven Knowledge Graph to Text Generation

Y. Song and C. Gardent. EMNLP 2025.

Improving Generalisation to OOD data

- We use our KG/Text alignment metric to create preference data
- DPO outperforms instruction tuning and 3 shot prompting on Out Of Domain Data



Multilingual Verbalisation of Knowledge Graphs

9 Languages

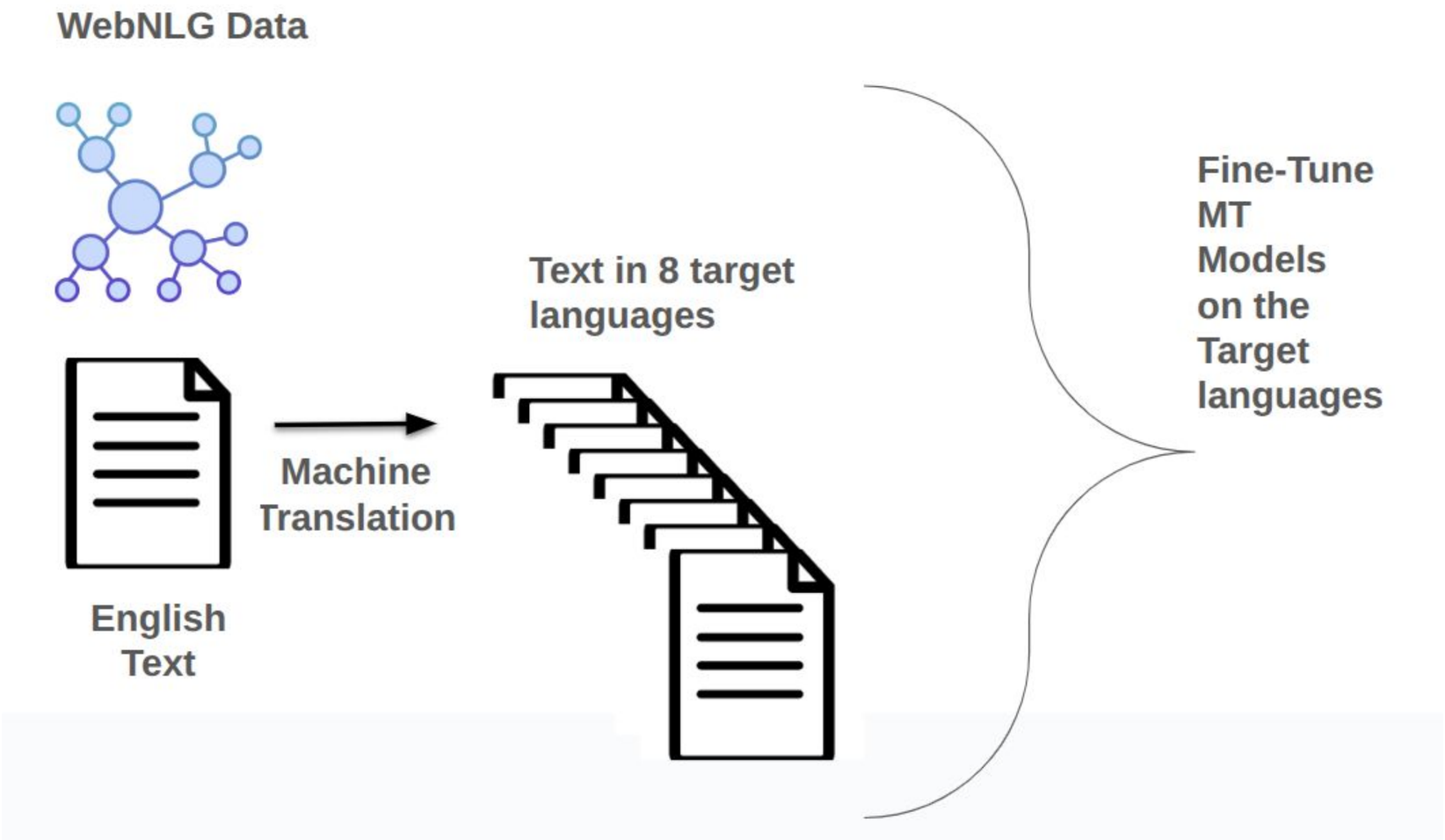
- HRL: Chinese, French, Russian, Spanish, English
- LRL: Breton, Irish, Maltese, Welsh

3 Methods

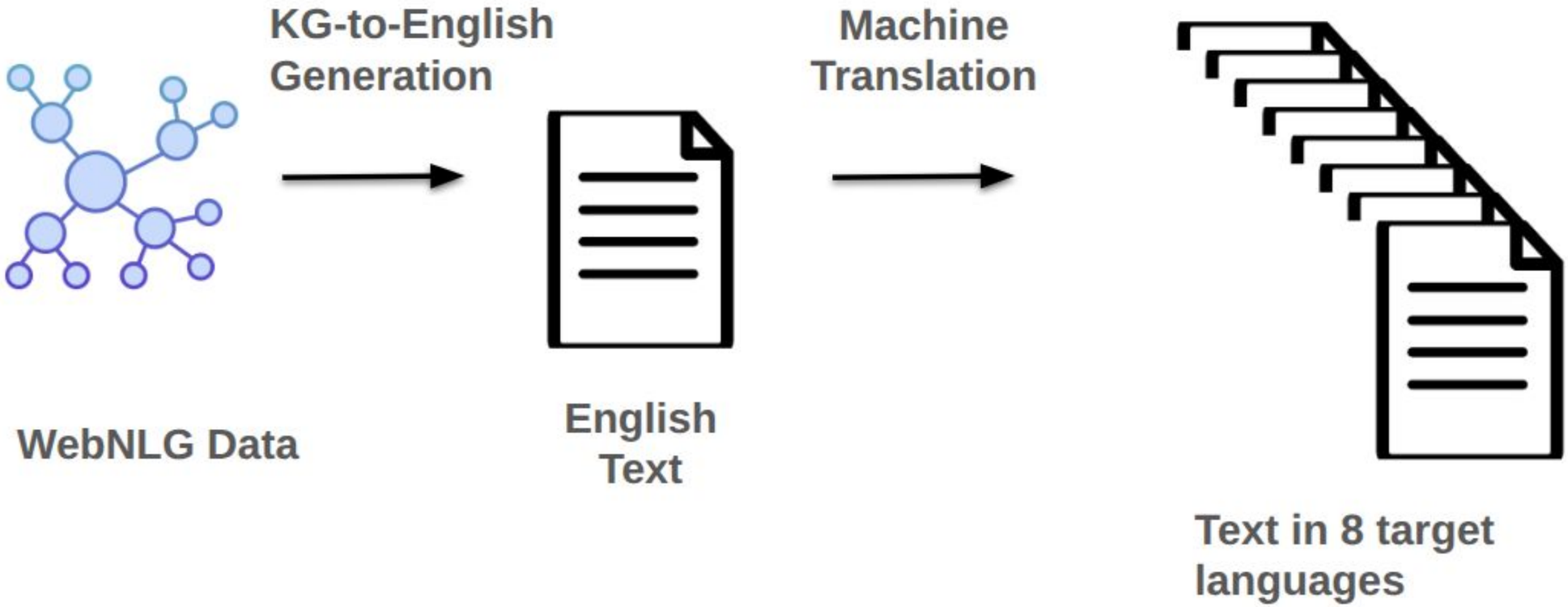
- Fine-tune Machine Translation Models on (Knowledge Graph, Text) pairs where the text is in one of the 8 target languages
- NLG-MT: Generate into English and machine translate into the target languages
- In-Context Learning: LLM Prompting

Y. Song, W. Soto-Martinez, A. Nikiforovskaya, E. Chapple and C. Gardent. Findings of EMNLP 2025.

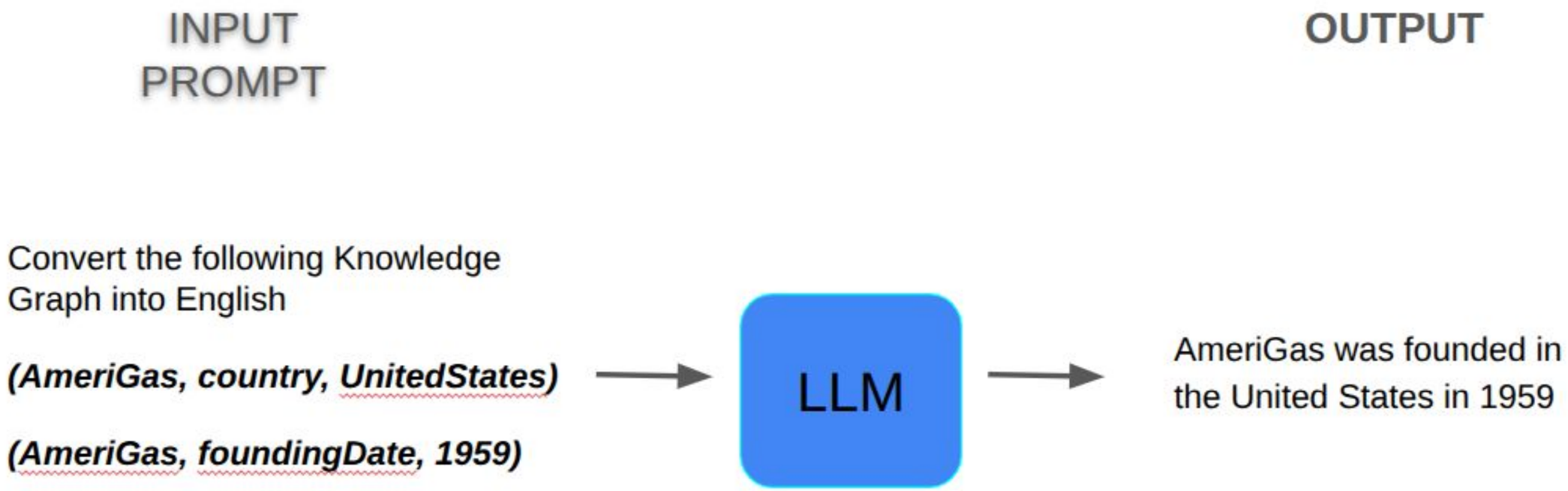
Fine-Tuning Machine Translation Models



Generating into English and Translating



In Context Learning



Performance and Generalisation

The **best prompt** contains target labels for entities and properties and few shots in the target languages such that the graphs are maximally similar with the input graph.

Prompting outperforms the other two methods

- On **In-Domain data**, the improvement mirrors the language distribution of the LLMs' pre-training corpora. The LLM has a strong advantage on languages it was well-trained on (English, Chinese). In languages less represented in the LLM’s training data, the gap between FewShot and the other methods is smaller.
- On **Out-of Domain data**, the improvement is largest for Low Resource Languages, demonstrating the ability of LLMs to generate multilingual text even for a task with minimal or no training data

IA Grand Est
ENACT

Collaborations

Research

- Albert Gatt, Utrecht University, The Netherlands. Co-supervision Yifei Song's PhD thesis
- Kelvin Han, Singapour. Question Under Discussion (QUD) as a means to reduce and assess LLMs hallucinations in Text-to-Text generation.

Interdisciplinarity

- Violette Abergel, CR CNRS Lyon. Models and Simulation for Patrimonial Data. Extracting temporal and spatial information form the Notre-Dame de Paris data science project.

Industry

- Cyber4Care, Nancy Start-Up
  - LLM4CARE, an AI assistant dedicated to business continuity,
- URS, Paris Start-Up
  - Preserving and discovering Knowledge

Teaching

Master TAL IDMC Nancy

- Data Science for Natural Language Processing (M1, 20h)
- Prompt Engineering (M2, 20h)
- Neural Approaches to Text Generation (M2, 20h)

Practicals: Yifei Song, Kun Zhang

IA Grand Est
ENACT

Thank you for your attention!

Questions ?

IA Grand Est
ENACT