Traitement Automatique des Langues au LORIA

LLMs, Apprentissage et Applications

Claire Gardent
CNRS / LORIA

MosAIk

28 permanent staff

- 4 professors, 22 associate professors, 1 senior research scientist (DR CNRS) and 1 research scientist (CR CNRS)

43 non permanent

- PhD candidates, engineers, alternants

KNOWLEDGE REPRESENTATIONS AND REASONING

This group specialises in AI systems which are able to manipulate explicit knowledge representations.

HYBRID AI

This group specialises in hybrid AI systems combining symbolic and statistic techniques.

FRUGAL AI

This group specialises in AI systems which rely on limited resources (data, storage and/or computing power).

NLP & LLM

This group specialises in AI systems used in Natural Language Processing with a focus on Large Language Models.

EXPLAINABLE AI

This group specialises in AI systems which exhibit some degree of interpretability and / or explainability.

Outline

Christophe Cerisara (CR CNRS/LORIA)

- Mille Pensées, an LLM that reasons in French
- Less back propagation to reduce computational costs

Yannick Parmentier (MCF UL/LORIA)

- Generating Grammar Exercises using a Hybrid neuro-symbolic approach

Gael Guibon (MCF Sorbonne Paris Nord/LIPN)

- Conversational Analysis with Language Models

Claire Gardent (DR CNRS/LORIA)

- Semantic Faithfulness in Conditional Text Generation

Mille-Pensées

Gabriel Lauzzana, Imane Ouada, Christophe Cerisara

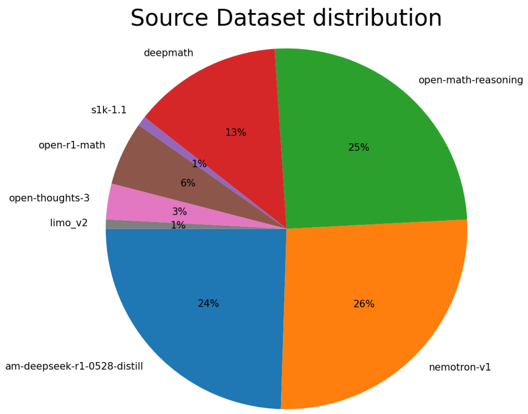
Reasoning LMMs reason in English

- Reasoning LLMs (o1, deepseek-R1 ...) can solve math questions
- However most reason in English
- Reasoning in French
  - Mistral Magistral (24B)
  - GPT-OSS (20B) fine-tuned on French: often reasons in English

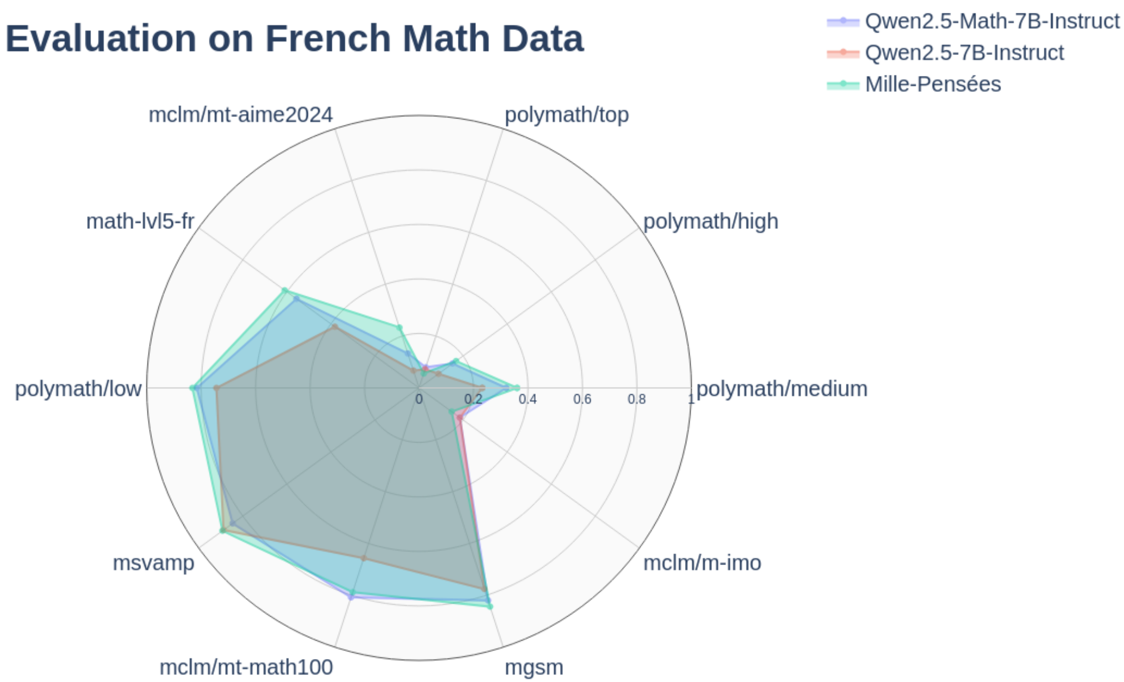
Goal: Post-train a small size LLM that reasons in French

Mille-Pensées - a Reasoning LLM that reasons in French

- Post-trained a non-reasoning LLM (Qwen2.5-instruct-7B) on FR (50%) and EN (50%) Math CoT: 1.6b tokens
- Training dataset - math problems translated from English
- Best LLM 7b that reasons in French
- Available on HuggingFace: https://huggingface.co/GLauzza/Mille-Pensees

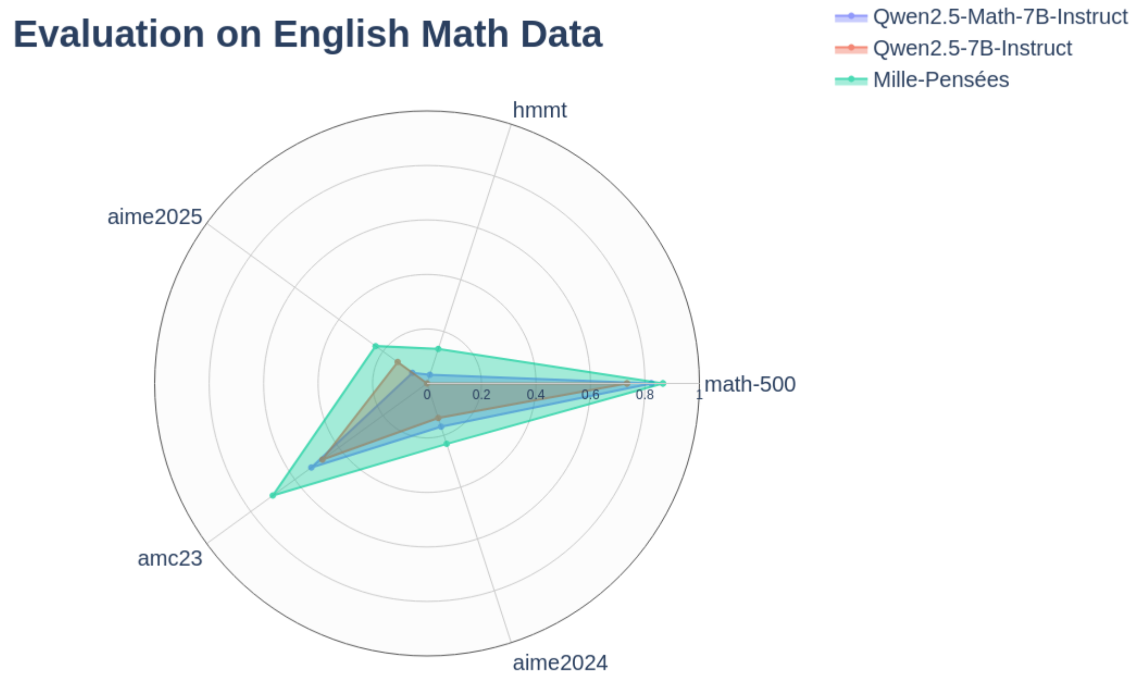


Reasoning in FR improves performance on Maths-FR



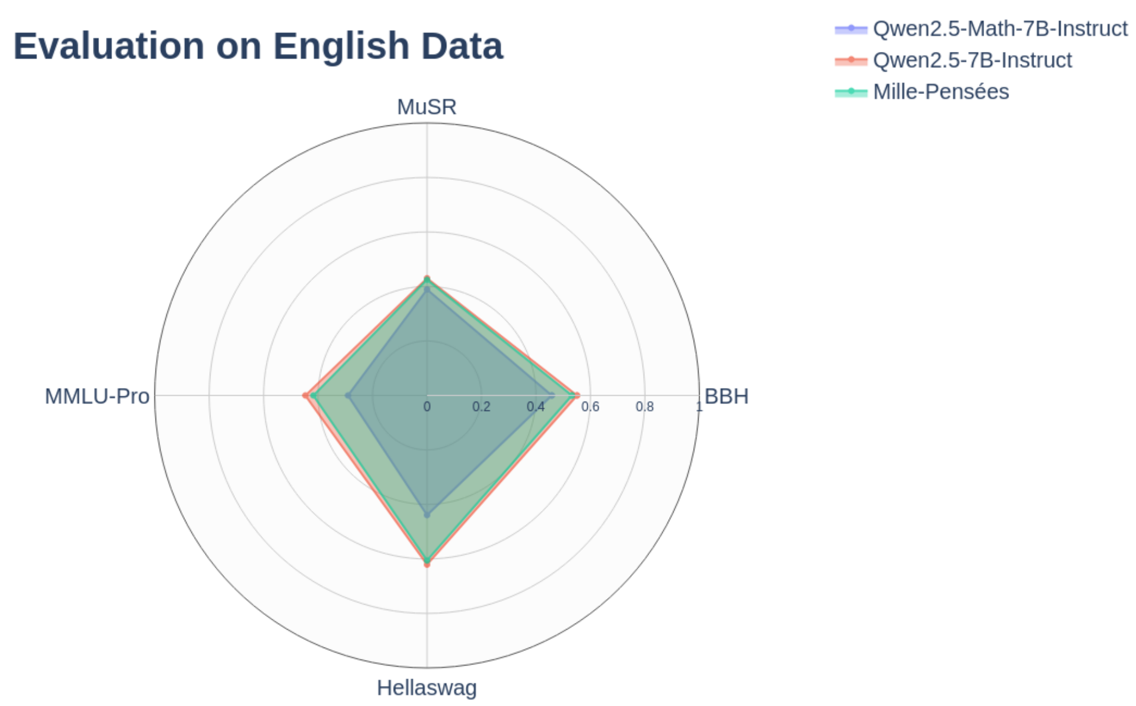
- Qwen-7B-FR >> Qwen-Math-7B-EN
- Reasoning in FR does not penalize maths solving capabilities

Also improves the performance for Maths-EN!



Our math post-training procedure/data-mix outperforms Alibaba's post-training pipeline.

Retains generic EN capabilities



MMLU: factual knowledge. BBH: NLP and Reasoning. Hellaswag: Commonsense Reasoning. MuSR: Multistep Reasoning

Back-propagation Reduction adaptation

Estelle Zheng, Christophe Cerisara

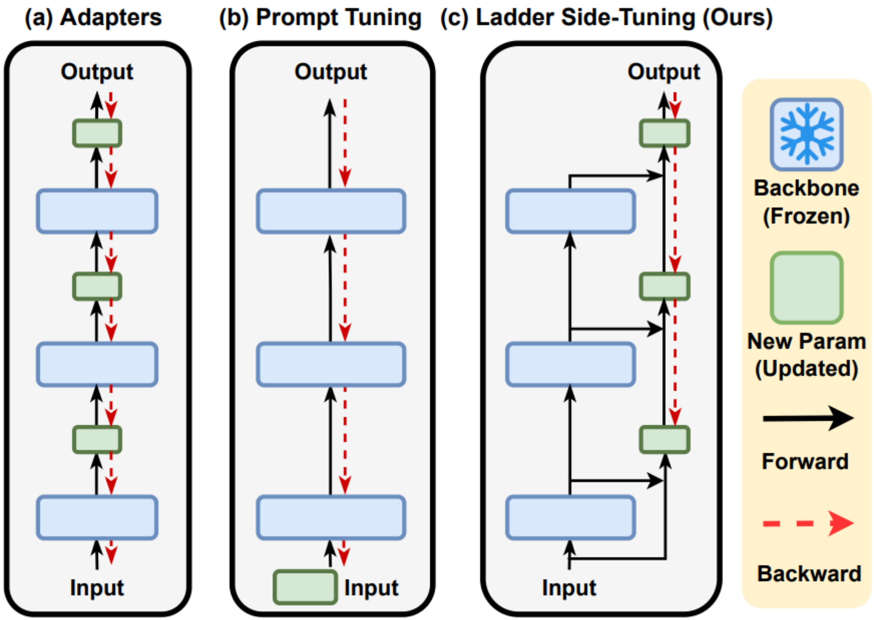
Reducing Training Costs

Training an LLM with N parameters

	FLOPs per token	**Memory**
Forward	$O(2N)$	$N/2$
Backward	$O(4N)$	$11N$

Goal: Reduce costs by reducing back-propagation

Ladder side-tuning



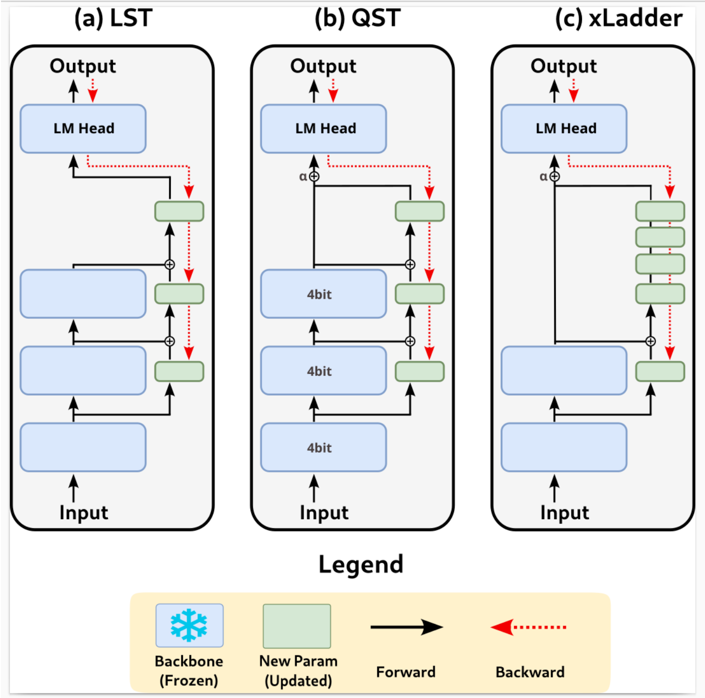
xLadder

Extends the ladder depth

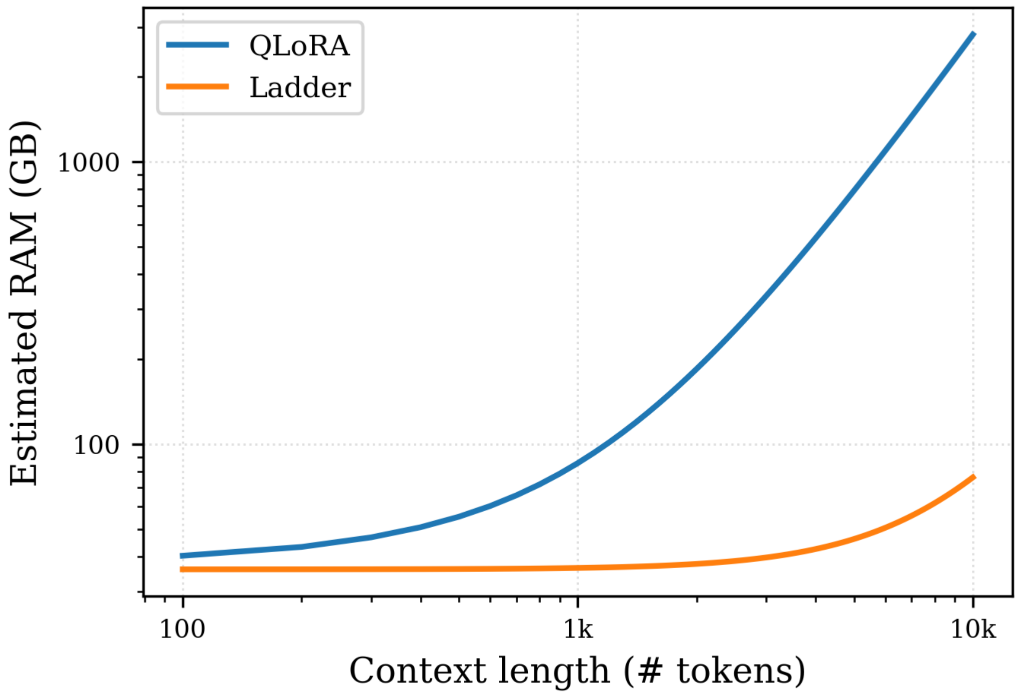
- this increases the nb of layers information go through during the forward path (which is equivalent to increase the nb of reasoning steps)

without increasing the nb of layers the backward pass go through

- this prevents vanishing gradient

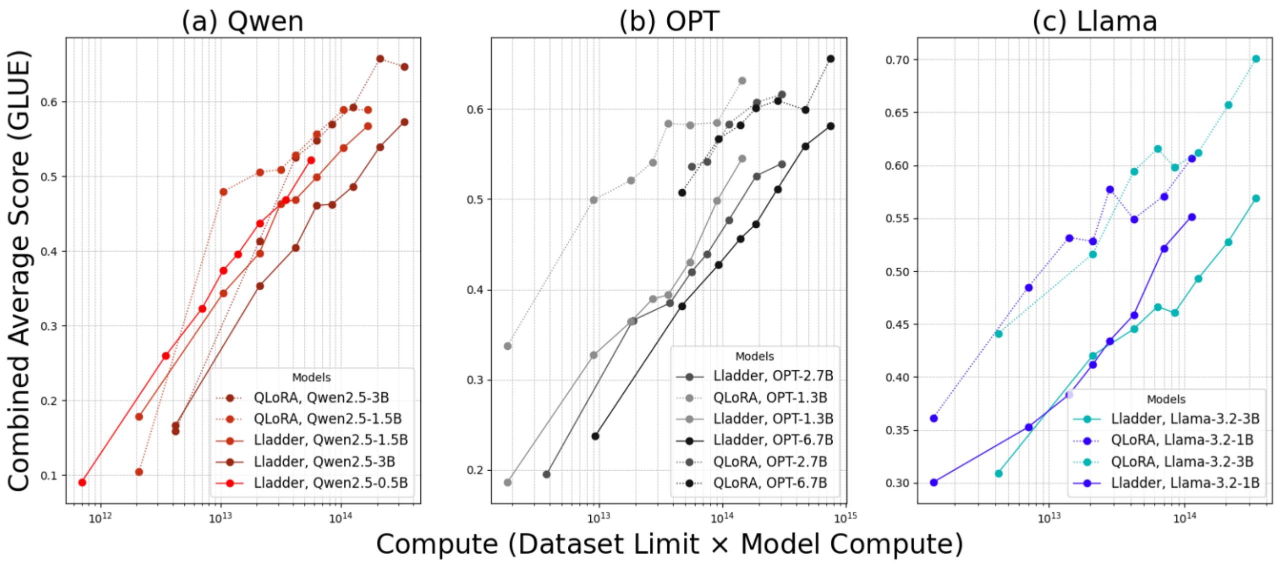


xLadder uses less memory



As the input length increases, xladder uses less memory than qLoRA

xLadder preserves performance on GLUE



On GLUE, xLadder performs on par with qLora at constant compute

xLadder preserves performance on Reasoning Benchmarks

xLadder performs on par with qLora on reasoning benchmarks

Model	Pure SFT?	MATH-500	AIME24	AIME25	AMC23	Minerva	Olympiad Bench
			Based on Qwen2.5-7B				
Qwen Base (Qwen et al., 2025)	✓	61.2 ± 0.4	8.7 ± 2.4	7.9 ± 4.3	32.5 ± 1.5	16.9 ± 1.2	30.2 ± 2.3
Qwen Instruct (Qwen et al., 2025)	✗	75.2 ± 0.5	8.7 ±2.4	8.7 ±4.3	38.5 ± 1.4	35.8 ±1.0	38.7 ±1.0
s1.1-7B (∗) (Muennighoff et al., 2025)	✗	80.8 ± 0.6	19.0 ± 3.2	21.0 ± 5.5	59.5 ± 3.7	37.5 ± 1.1	48.2 ±1.4
Full SFT (∗) (Wang et al., 2025)	✗	58.6	10.0	7.1	45.3	24.6	27.6
QLoRA	✓	**70.4** ± 1.4	**8.7** ± 1.8	6.0 ± 4.3	41.5 ± 1.4	**33.5** ± 1.3	32.0 ± 0.8
Ladder (our)	✓	68.9 ± 1.6	8.0 ± 1.8	**8.0** ± 1.8	**47.5** ± 7.3	29.3 ± 1.2	**34.3** ± 0.5

Apprentissage des langues assisté par ordinateur - le projet GramEx

Une méthode hybride pour la génération d'exercices

Yannick Parmentier

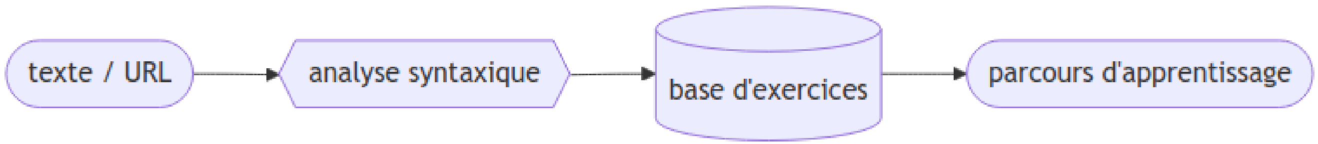
Le Projet GramEx

Objectifs et Contraintes

- génération automatique d' ***exercices ciblés*** (notion grammaticale, niveau apprenant)
- à partir de textes
- nécessité de ***garantir la correction*** des exercices générés

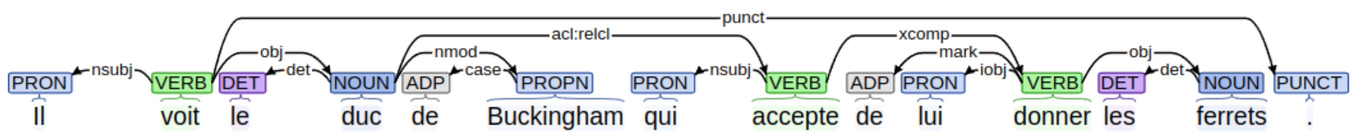
Méthode hybride pour la génération automatique d'exercices

Analyse syntaxique neuronale + Transformation symbolique



Exemple

Texte : *Il voit le duc de Buckingham qui accepte de lui donner les ferrets.*
Analyse syntaxique :



Exercice : Dans la phrase ci-dessous, sélectionnez le sujet du verbe souligné ...

Questionnements scientifiques

Filtrer automatiquement les textes en fonction du niveau des apprenants

- Affinage d'un grand modèle de langue du français (CamemBERT) sur un corpus de livres pour enfants créé semi-automatiquement (Van Ngo et Parmentier, 2023)

Prédire la *difficulté* d'un exercice de grammaire

- Constitution d'un corpus d'exercices annotés en niveau CECRL (A1/A2/B1/B2/C1/C2) à partir de ressources en ligne issues de la discipline Français Langue Etrangère (FLE)
- Application d'algorithmes d'apprentissage automatique

Questionnements scientifiques (suite)

Estimer la *fiabilité* des annotations produites par un analyseur syntaxique sur des données hors-domaine

- Constitution d'un corpus combinant analyses automatiques et références humaines (gold-standard) pour entraîner un modèle de fouille d'erreurs (Toussaint et a l., 2024)

Semantic Faithfulness in Conditional Text Generation

Claire Gardent

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 $\rightarrow$ Text

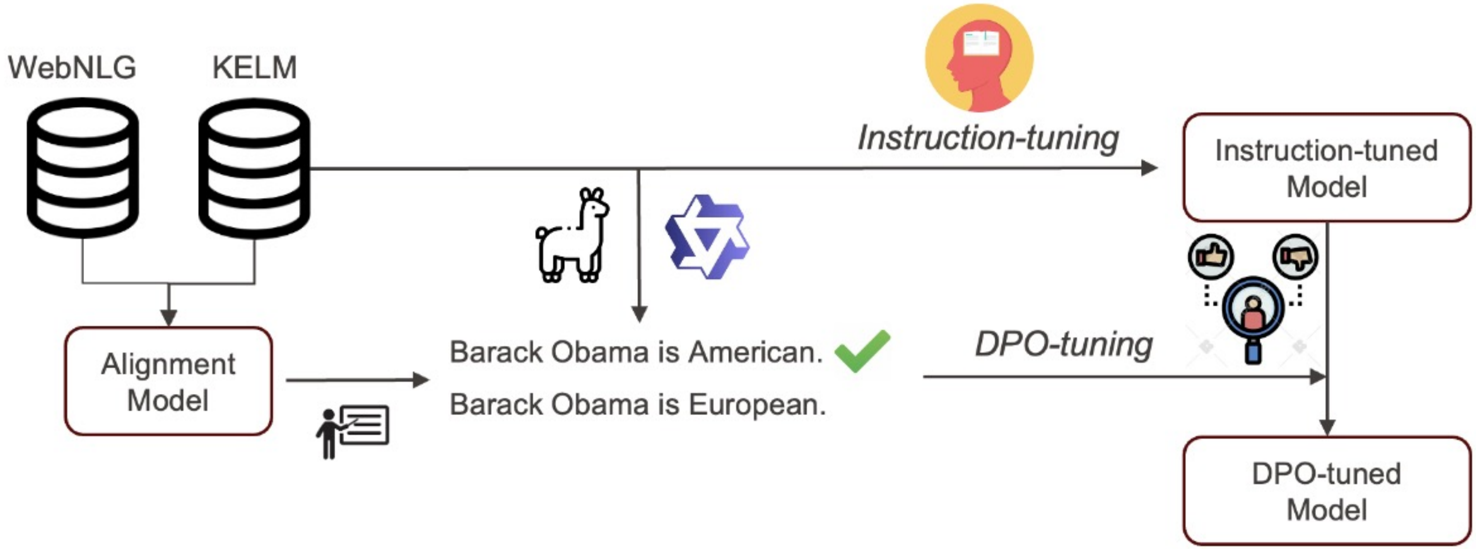
- Multilingual Verbalisation of Structured Data.
- Improving Generalisation and Faithfulness
- Evaluating Faithfulness

Text $\rightarrow$ Text Cross-lingual Biography Generation. Reducing Data and Model Biases

Analysis Text $\rightarrow$ Data

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

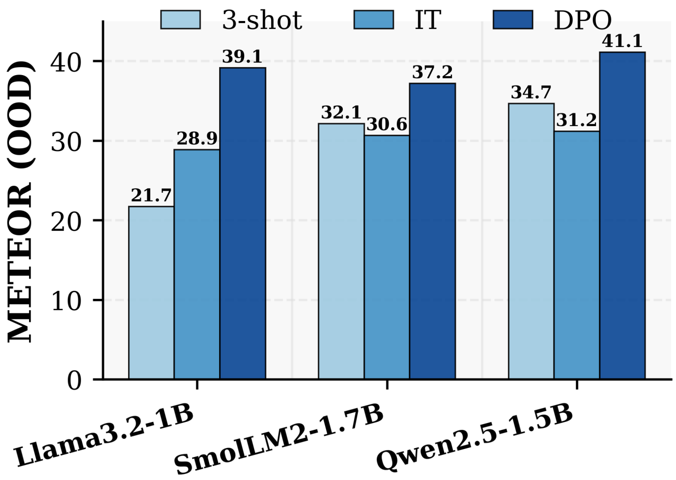
MuCAL: Contrastive Alignement for Preference Driven Knowledge Graph to Text Generation



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 Graph

9 Languages

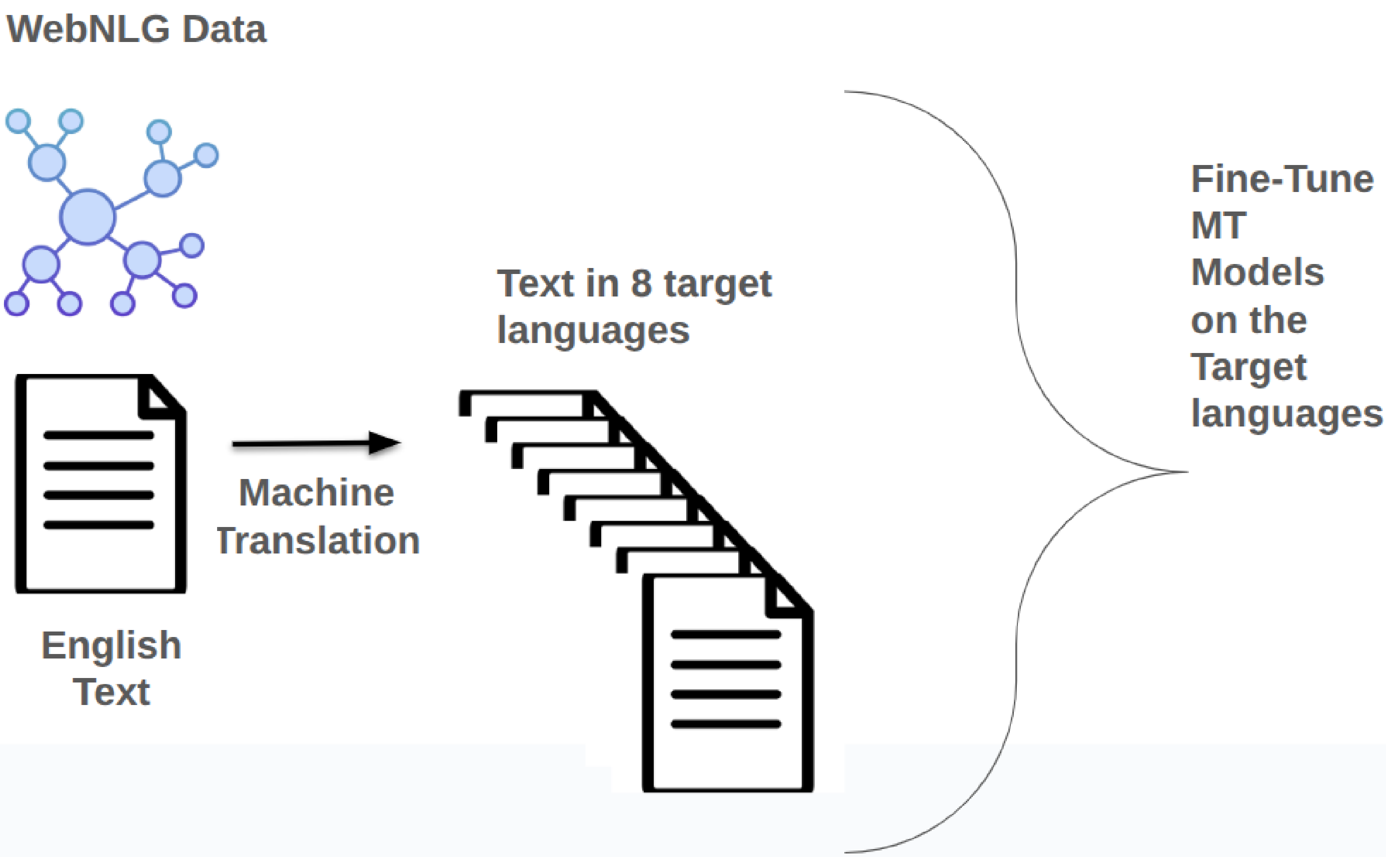
- HRL: Chinese, French, Russian, Spanish, English
- LRL: Breton, Irish, Maltese, Welsh

3 Methods

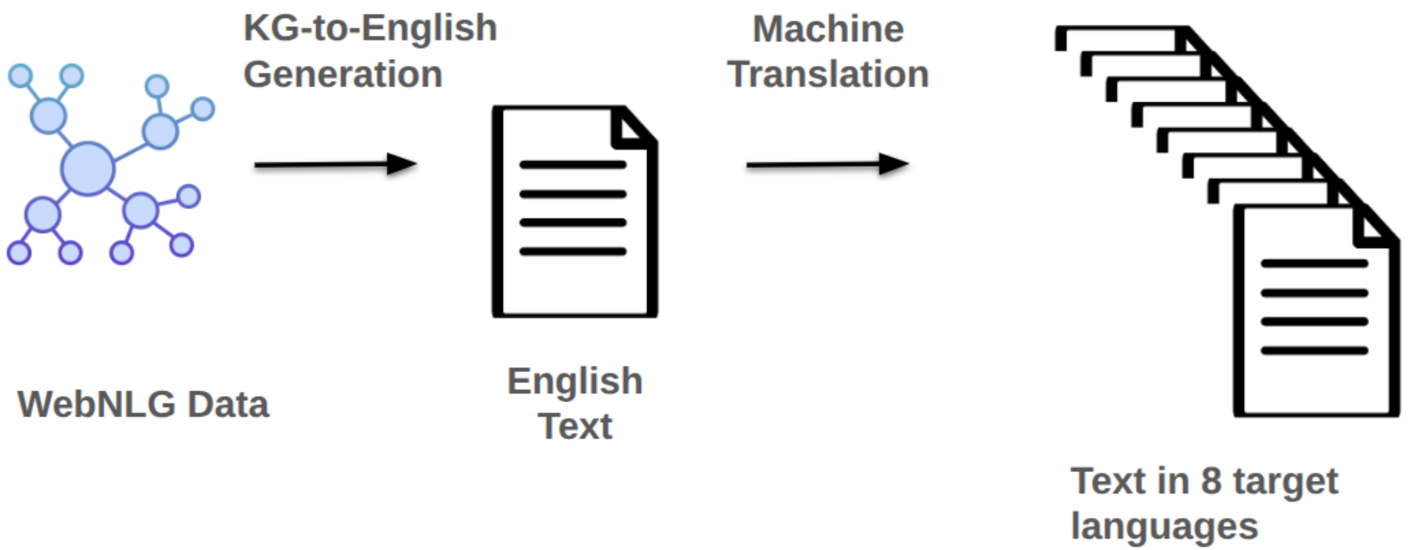
- Fine-tune Machine Translation Models on (Knowledge Graph, Text) pairs where the text is in one of the 8 target languages
- NLG-MT: Generate into English and machine translate into the target languages
- In-Context Learning: LLM Prompting

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

INPUT PROMPT

Convert the following Knowledge Graph into English

(AmeriGas, country, UnitedStates)

(AmeriGas, foundingDate, 1959)

OUTPUT

AmeriGas was founded in the United States in 1959

Performance and Generalisation

The best prompt contains

- target labels for entities and properties and
- few shots in the target languages whose graphs are maximally similar with the input graph.

Prompting outperforms the other two methods

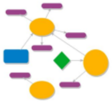
- On **In-Domain data**, the improvement mirrors the language distribution of the LLMs' pre-training corpora. The LLM has a strong advantage on languages it was well-trained on (English, Chinese). In languages less represented in the LLM’s training data, the gap between FewShot and the other methods 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

Conversational Analysis

Gael Guibon

Conversational Analysis with Language Models

Controlling the Conversation with Dedicated Ontologies



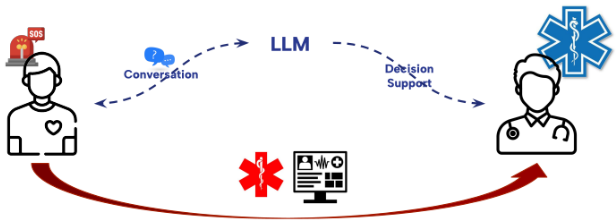
Part I - Define Conversational Control Ontology Design

RQ1: How can an ontology be used to fine-tune LLMs for controlled conversational guidance?

Part II - Enable Conversational Control Fine-Tuning Approaches

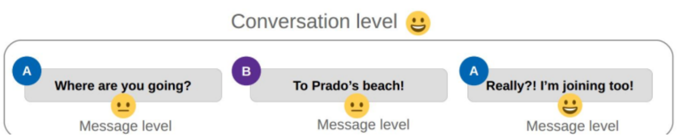
RQ2: Can the LLM compliance to ontologically-defined conversation aspects be enabled through fine-tuning?

Adaptation and Use of LLMs for Medical Emergency Regulation

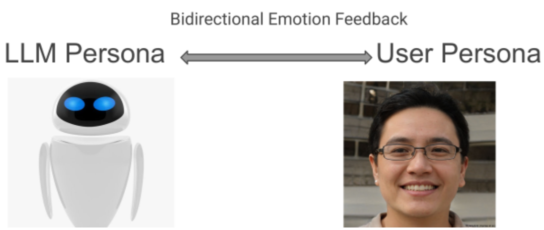


Emotion Recognition in Conversations

- Consider the **conversational context**
- **Identify** emotional messages among the neutral ones
- **Assign** the correct emotion to identified messages



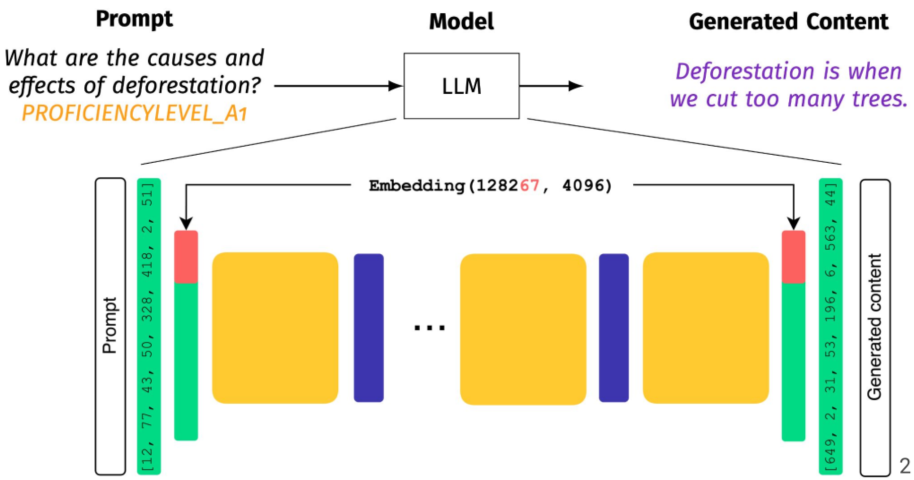
Personalization of Emotional Feedback through Emotion Recommendation



Conversational Analysis with Language Models

Controlling the conversation by controlling the LLM output and aligning strategies

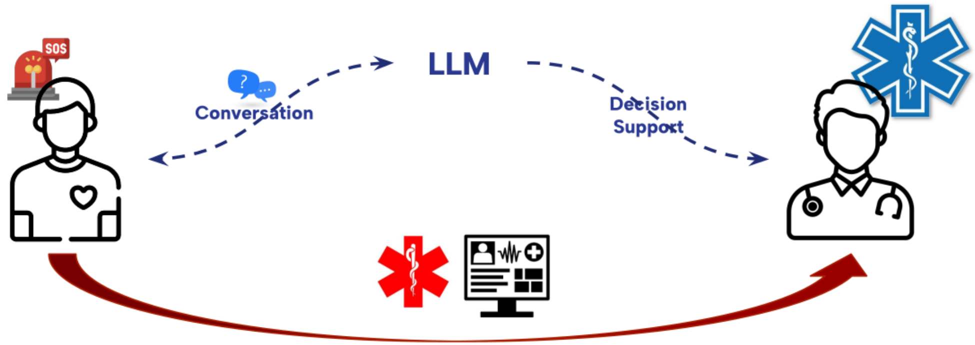
- **RQ1:** How can we use an ontology to fine-tune LLMs for controlled conversational guidance?
- **RQ2:** Can we enable the LLM compliance to ontologically-defined conversation aspects through fine-tuning?



Gendron, Gaël Guibon, Mathieu D'aquin. Towards Ontology-Based Descriptions of Conversations with Qualitatively-Defined Concepts. In Proceedings of the 19th International Conference TOTh – Terminology & Ontology: Theories and applications, Chambéry (France), June 2025. https://hal.science/hal-05240495v1?

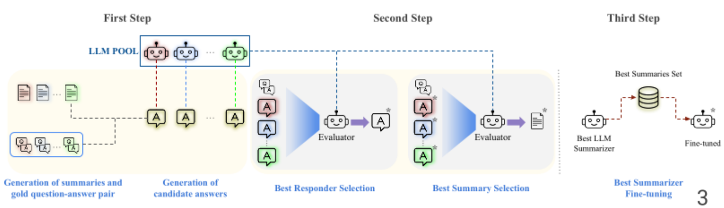
Conversational Analysis with Language Models

Adaptation and Use of LLMs for Medical Emergency Regulation



1. **Retrieve** critical patient information **as quickly as possible** in order to dispatch the appropriate emergency teams.

2. **Predict** the severity level (e.g., determine whether the situation is a life-threatening emergency).



3

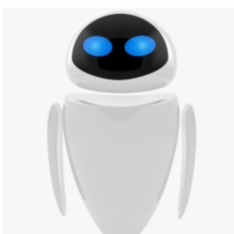
Mohamed Imed Eddine Ghebriout, Gaël Guibon, Ivan Lerner, and Emmanuel Vincent. 2025. QUARTZ: QA-based Unsupervised Abstractive Refinement for Task-oriented Dialogue Summarization. In Findings of the Association for Computational Linguistics: EMNLP 2025, pages 14689–14706, Suzhou, China. Association for Computational Linguistics.

Conversational Analysis with Language Models

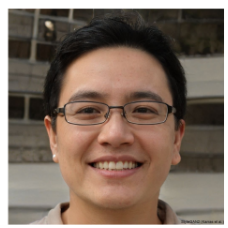
Personalization of emotional feedback through emotion recommendation

We **fine tune LLM** through **recommendation system**, to create a **conversational agent** with **persona** that suits well the **emotional needs** of the users.

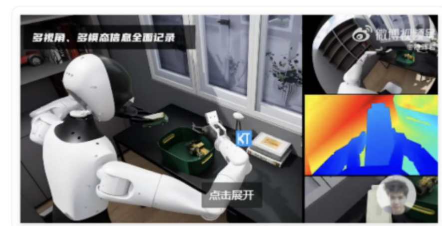
Persona



Persona



Interactive intelligence is another highlight of the Lingxi X2. The robot is equipped with a customized multi-modal interaction large model, supporting various interaction methods such as voice and vision. The emotion computing engine allows it to understand human emotions and respond accordingly, significantly enhancing the naturalness and immersion of human-robot interaction.



Conversational Analysis with Language Models

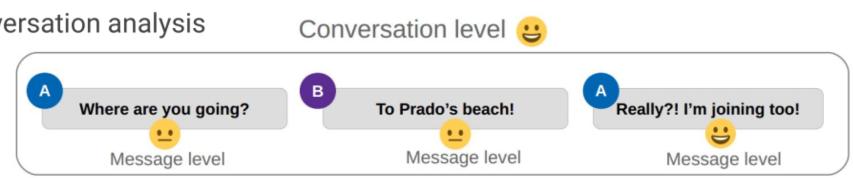
Emotion Recognition in Conversations

Emotion in Conversations

- **Fine-Grained Emotions and Emojis**
  - Categorization, recommandation, prediction and intensity
  - Irony and humor detection
  - Emoji Personalized Recommendation
- **Conversational Context**
  - Optimized context representation
    - By flow, speaker, turn, etc.
  - User Generated Content
  - Nested conversation analysis

Real Data and Training Adaptation

- **Data-Adapted Approaches**
  - Few-Shot Learning (not prompt-based)
  - Meta Learning
- **Model Adaptation**
  - Domain Adaptation for Language Models
    - Which control level ?
  - Two-level conversational context



Guibon, Gaël, et al. "An adaptive layer to leverage both domain and task specific information from scarce data." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37. No. 6. 2023.