Controlled Conversational Models through Conversation-Dedicated Ontology

Barbara Gendron, under the supervision of Mathieu d'Aquin and Gaël Guibon

First-year PhD student at LORIA, CNRS, Université de Lorraine (Nancy, France)

barbara.gendron@loria.fr | b-gendron.github.io | b-gendron | barbara.gendron

1. Context

Recent advances in Large Language Models (LLMs) have improved conversational agents' realism and compliance towards human requirements and needs. However, **controlling conversation flow towards positive outcomes remains crucial**.

This Ph.D. aims to **represent conversational knowledge using an ontology to enable language model control**. Ontologies allow to model the knowledge in a domain, defining concepts and characterizing relations between them. While often used for domain-specific knowledge, few have explored using ontologies to guide conversation flow. Convology[1] is a recent example focusing on managing health conversations. We plan to extend Convology's conceptualization capacities to a more general setup, therefore adaptable to general-purpose user/agent conversations.

Want to Know More?

An example: OntoGPT for Readability Level Assessment (seminar slides)

2. Methodology

Tools. Protégé, Pellet, owlready2, PyTorch, Huggingface transformers and parameter-efficient fine-tuning libraries, LoRA adapters.

Ontology. Progressively incorporate and infer on linguistic features such as part-of-speech tags, affective computing such as emotions or dialog acts.

Experimental Setup. Toy example design process is aligned with LLM abilities. So far, we considered readability level and sentence length.

Challenges. It is not straightforward that the knowledge the ontology brings can be accurately learnt and applied by a language model, whether it be decoder-only or encoder-decoder.

Conversation Control. Decisions to use certain outputs rather that other will be directly associated to the ontological dimensions of the current conversation that have driven those choices, which makes the difference with black-box models. In the end, this could help to discourage the generation of harmful content, this bringing controllable ethics to human-machine conversations.

[1] Dragoni et al. *Web Semantics*. 2021.

3. Motivation & Objectives

The objective of this thesis is to develop **knowledge-enhanced conversational models** that exploit Large Language Models (LLMs) and Ontologies. This consists in improving state-of-the-art LLMs by providing **structured knowledge** to open-domain conversational agents.

Objectives:

- Build a conversation ontology that accounts for interpersonal relationships concepts and their evolution.
- Integrate and assess ontology understanding during fine-tuning.
- Bring control on conversational LLM outputs through encapsulated conversation knowledge.

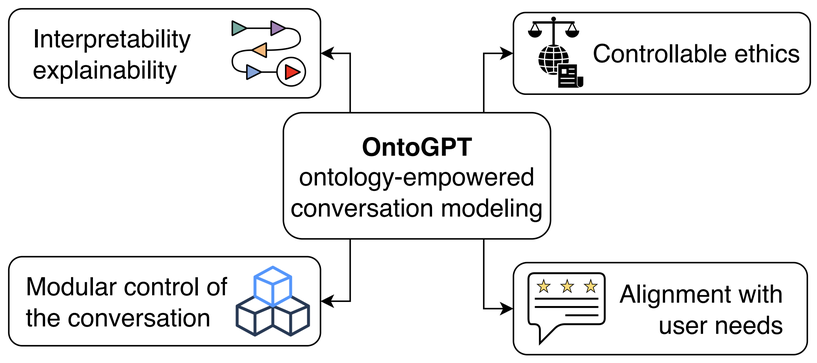


Figure 1: The benefits of ontology-LLM hybridation systems for conversation modeling

4. OntoGPT: LLM Fine-Tuning Based on Ontology Validation

OntoGPT fine-tunes LLMs using LoRA adapters. It aims at improving generation by learning a classification task guided by the ontology knowledge.

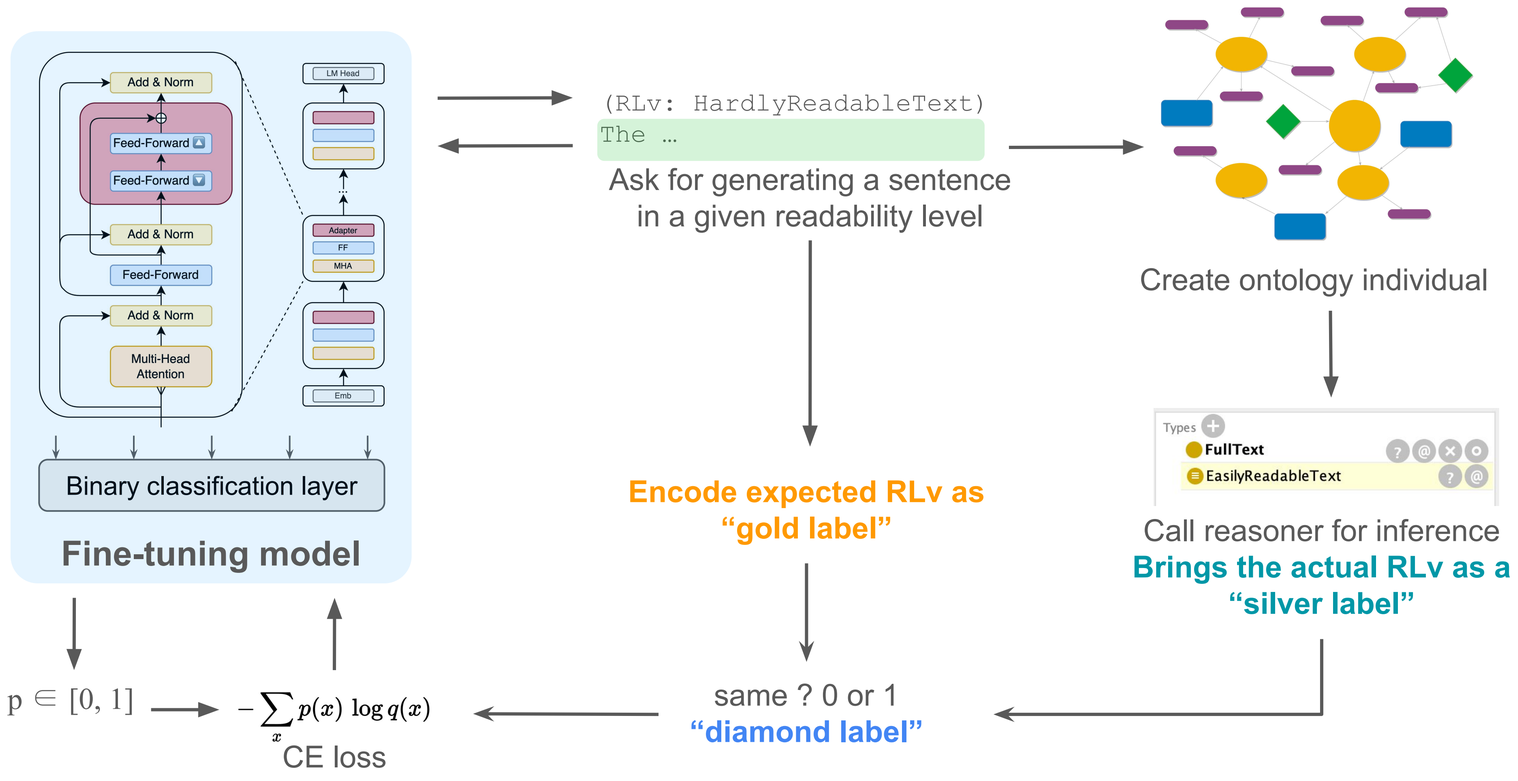
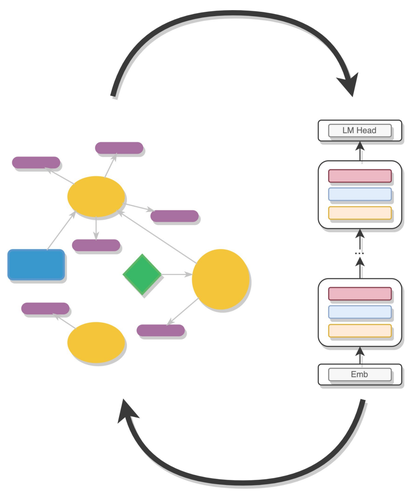


Figure 2: OntoGPT is an end-to-end integration pipeline where the ontology information is assimilated at fine-tuning time. This example focuses on readability level assessment task.

5. Advances & Perspectives

Ontology Engineering:

- Semantic Web basics (RDF, RDFS, OWL), ontology engineering methodologies and tools (Protégé).
- Hybridization of LLMs and ontologies in a fine-tuning pipeline.
- Automatic creation and management of ontology individuals to fine-tune through ontology validation.



Language Models:

- Build a lightweight decoder-only language model from scratch to testbench pre-training and fine-tuning scenarios.
- Develop generation-based fine-tuning that relies on a classification task evaluated by an ontology. Involves challenges in stability, knowledge transfer and computational efficiency.