Large language models and foundational ontologies

Paweł GARBACZ [1]
The John Paul II Catholic University of Lublin
ORCiD ID: Paweł Garbacz https://orcid.org/0000-0003-4145-7528

Abstract. Large language models are capable of translating natural language texts into context-free grammar languages. The paper presents an initial assessment of whether such models can be used to produce ontological theories that formalise natural language descriptions of certain situations. More specifically speaking, I will focus here on translating a small set of natural language descriptions of some situations of ontological interests into a fixed formal ontological framework. The model I use will not be trained or fine-tuned for this purpose but prompted. In order to build the appropriate prompts I will take advantage of the formalisations from the 17th volume of the Applied Ontology journal, where six examples of such situations were formalised within the context of seven upper-level formal ontologies.

Keywords. large language model, formalisation, formal ontology, CLIF

Introduction

Recently, we have witnessed that large language models, e.g., GPT models issued by OpenAI, are capable of solving a great deal of tasks in computational linguistics. In particular, they are known to be able to generate efficient computer scripts from functional requirements specified in natural language – see, for instance, the details of the Copilot project. It is also known that they can be used to translate natural language questions into such formal languages as SPARQL (see, for example, [1]).

In the context of formal ontology, these successes may raise the question of whether a large language model can be trained to generate ontological theories that formalise natural language descriptions of certain situations. More generally, one can ask whether we can train them to translate natural language texts into formal theories.

Since this is one of the early attempts in this respect, I will focus on a relatively modest goal of translating a small set of natural language descriptions of some situations of ontological interests into a formal ontological framework. The large language model (LLM) will not be trained or fine-tuned for this purpose but simply *prompted*, so I will test its 'out-of-the-box' capability of discovering certain latent mappings between natural and formal language in the context of formal ontology. For the purpose of this paper, I will use the formalisations from the 17th volume of the Applied Ontology journal, where seven main upper-level ontologies were presented, and each ontology was tasked

[1] Corresponding Author: Paweł Garbacz, garbacz@kul.pl

to represent six examples of situations of ontological interest as seen from its perspective - for the details, see section 3 below. The examples will be referred to as the AO-examples and will be identified in the same way as in the journal itself, i.e., as Case 1, Case 2, etc.

Looking at this research experiment from the perspective of the programme of weak artificial intelligence in the sense of J. Searle, one can say that if an LLM can translate a sufficient number of natural language descriptions in accordance with an ontological framework, i.e., taking into account the posits of a formal ontology, then the LLM can be thought of as being able to understand this ontology at least in the sense of being able to simulate such understanding. Note that the editors of the volume in question made it clear that such applications of foundational ontologies facilitate our understanding thereof. Still, I hope the research presented below can be appreciated outside also this programme.

As the reader will shortly realise the research presented in this paper exhibits many loose ends, both in terms of the theoretical depth and in terms of the technical robustness. In some respects the paper is more like a position paper - however some of its drawbacks seem inevitable as they are due to the relatively novel research tasks it attempts to achieve.

1. Related Work

The most relevant, indeed a trigger for the current paper, is the research reported in [2]. The authors introduce there a couple of LLM models from the LogicLLama family of models, which are fine-tuned adaptations of the open-source model LLaMA-7B with the data from the MALLS dataset – all these artefacts are available from `https://huggingface.co/yuan-yang`. The MALLS dataset extends the FOLIO dataset, which contains approx. 2000 human-made translations from natural language to first-order logic formalism - see [3]. The former dataset, which consists of approx. 34k of natural-language-to-first-order-logic translations, was created by prompting the GPT-4 model.

In order to see whether these models are capable of simulating proper ontological analysis, consider the following test: each AO-example is to be translated to a first-order logic formula by the LogicLLaMA-7b-direct-translate-delta-v0.1 model from the LogicLLama family of models. The outcomes of this process are shown in table 1.

As the reader may easily verify, out of these six examples no translation comes even close to the actual meaning of the respective natural language text. The most obvious errors involve:

1. the model seems to confuse the existential and universal quantifiers;
2. the model does not properly distinguish different entities and uses the same variable to refer to them;
3. the selection of predicates is sometimes haphazard;
4. one translation is not even syntactically well-formed.

One may *guess* that the reason for this failure is the fact that the MALLS dataset contains single sentences, some of which are syllogistic-like generalisations like 'A movie that is a comedy can make people laugh.', 'If a person is a scientist, they either work in a laboratory or a research institution.' or 'All mammals have hair or fur on their bodies'.

AO-example	FOL formula
[Case 1] There is a four-legged table made of wood. Some time later, a leg of the table is replaced. Even later, the table is demolished so it ceases to exist although the wood is still there after the demolition	$\exists x(Table(x) \land FourLegged(x) \land MadeOfWood(x) \land \exists y(Leg(y) \land PartOf(y,x) \land Replaced(y)) \land \exists z(Demolished(z) \land Table(z) \land PartOf(z,x) \land CeasesToExist(z) \land WoodStillThereAfterDemolition(x)))$
[Case 2] Mr. Potter is the teacher of class 2C at Shapism School and resigns at the beginning of the spring break. After the spring break, Mrs. Bumblebee replaces Mr. Potter as the teacher of 2C. Also, student Mary left the class at the beginning of the break and a new student, John, joins in when the break ends.	$\forall x \forall y \forall z(Teacher(x) \land Class(y) \land School(z) \land Shapism(z) \land TeacherOf(x,y) \land ResignsAt(x,y,SpringBreak) \land Replaces(x,y,MrsBumblebee) \land Student(Mary) \land Joins(Mary,John) \rightarrow (LeavesAt(Mary,y,SpringBreak) \land JoinsAt($
[Case 3.1] A flower is red in the summer. As time passes, the color changes. In autumn the flower is brown.	$\forall x(Flower(x) \rightarrow ((Red(x) \land Summer(x)) \land \neg Red(x) \land Autumn(x) \land Brown(x)))$
[Case 3.2] A man is walking when suddenly he starts walking faster and then breaks into a run.	$\forall x(Man(x) \land Walking(x) \land SuddenlyWalksFaster(x) \rightarrow BreaksIntoRun(x))$
[Case 4] A man is walking to the station, but before he gets there, he turns around and goes home.	$\forall x(Man(x) \land WalkingToStation(x) \rightarrow (\exists y(Station(y) \land BeforeGettingTo(x,y)) \land TurnsAroundAndGoesHome(x)))$
[Case 5]Marriage is a contract between two people that is present in most social and cultural systems and it can change in major (e. g. gender constraints) and minor (e.g. marriage breaking procedures) aspects.	$\forall x \forall y \forall z(Marriage(x) \land Person(y) \land SocialOrCulturalSystem(z) \rightarrow (ContractBetween(y,x) \land PresentIn(x,z) \land (ChangeInMajorAspects(x,y,z) \lor ChangeInMinorAspects(x,y,z))))$

Table 1. LogicLLama translations

Other similar research concerns the capability of language models to perform deductive inferences ([4], [5]), to evaluate the validity of reasoning processes when they are materialised in the form of certain texts ([3], and to generate proofs ([6]).

[1] shows the benefits of using the so-called controlled natural languages (to be more specific, SQUALL and Sparklis) in the semantic parsing of natural language queries so that one can answer them against a knowledge graph using SPARQL.

A more general perspective on the current and future, possible dependencies between large language models and ontologies is outlined in [7]. The gist of the argumentation there is that it is likely that the former will be adapted to become useful tools for the development of the latter, but it is unlikely, or even impossible, that (i) the latter will become obsolete and that human ontologists will become obsolete agents in this development. The paper also mentions a number of recent attempts at using LLMs to generate OWL ontologies from short natural language texts. For the lack of space, let me refer to just one of them, namely [8]. This preprint describes an application, i.e., a Protege plugin, that uses a fine-tuned GPT-3 model to translate short, simple natural language sentences into the OWL Functional Syntax. Unfortunately, I was unable to get hold of the Protege plugin the paper describes, so I cannot test it against the AO-examples. The training and testing examples mentioned in the paper are single, simple sentences like 'Every rose is a flower', 'A mother is a female who has at least 1 child', or 'Cora and Meena hate each other'. Therefore, using the above lesson learnt from LogicLLama family of models, I would guess that the model from [8] might also be incapable of translating AO-examples.

2. Design

The goal of this paper is to check whether given (i) a certain situation described by a few-sentence long text in English and (ii) a formal foundational ontology, a large language model is able to translate this text into the formal language of the ontology in such a way that, if I may use Plato's metaphor from Phaedrus 265e, the translation carves nature at the joints delineated by the ontology.

After a number of less successful attempts, the overall strategy I follow in this paper has the following steps:

1. data ingestion:
   (a) selection of appropriate training pairs consisting of natural language description and its formal translation;
   (b) collection of appropriate natural language descriptions to test to what extent an LLM can be 'few-shot trained' to formalise such texts
2. preparation of the training prompts
3. querying the LLM with these prompts
4. evaluation of the LLM's responses

The LLM model of choice was GPT-4 ([9]). All queries sent to this model (on November 21st, 2023) were given the default parameters except for the so-called temperature, which was set to 0. All the input data, Python scripts, and data artefacts mentioned in this paper can be found in a GitHub repository.[2]

3. Realisation

3.1. Ontological prompts

One of the outstanding features of large language models is their capability of recognising certain syntactic patterns on the basis of very few data instances - that's why they are called few-shot learners [10]. In the context of this paper's goal, this feature is essential since the amount of available training data, i.e., natural language tests translated into formal ontological theories, is negligible if compared to other types of texts.[3]

For the purpose of this paper, I will peruse the formalisations from the 17th volume of the Applied Ontology journal, where the seven main upper-level ontologies are presented. Each ontology is used to formally represent six examples, i.e., AO-examples, where a full AO-example consists of

1. piece of short English text describing a certain situation of ontological interest
2. goal of this description in the context of formal ontology programme
3. focus, i.e., what particular aspects of the given situation we want to emphasise and which we will ignore.

For instance, the full Case 1 AO-example is quoted below:

> Case 1: (Section 3.1. Composition/constitution) "There is a four-legged table made of wood. Some time later, a leg of the table is replaced. Even later, the table is demolished so it ceases to exist although the wood is still there after the demolition." GOAL: The example aims to show if and how the ontology models materials, objects, and components and the relationships among them. FOCUS: The relationship between the wood and the table and the table's parts over time. (Artefacts and functions are not the focus.) [11, p. 8]

Since this is a relatively early stage of this type of research, I neglected the goal and focus components and used only the examples' texts. Also, given the space limitations for this paper, I used only one set of formalisations, i.e. the one based on the DOLCE ontology (see: [12, pp. 52-66].[4]

In order to obtain computer-friendly rendering of the DOLCE's formalisations of the AO-examples, I translated them into the CLIF language ([14]). Table 2 shows the details of one such formalisation – the rest can be found in the paper's repository.

---

[2] `https://github.com/mereolog/llmantogy.git`. This repository stores all scripts I used, so in principle the reader can reconstruct the finding reported here. Still, given the probabilistic nature of the model and its continuous development, it is unlikely that they will obtain exactly the same results. Also one needs to obtain the appropriate access to the GPT-4 model, i.e., the so-called API key, and pay for the queries. The cost of all queries reported in this paper was 1.39 USD.

[3] I assume here that description logic languages are not expressive enough for this purpose.

[4] Obviously, you can repeat this exercise using other ontologies presented in the aforementioned volume provided that the ontology you choose is indeed computational friendly – see more comments on this in [13].

Original FOL formula	CLIF sentence
$SocMarriage(x) \rightarrow C(x)$	`(if (SocialMarriage x) (Concept x))`
$LegMarriage(x) \rightarrow C(x)$	`(if (LegalMarriage x) (Concept x))`
$LegMarriage(x) \rightarrow C(x)$	`(if (SocialRelationship x) (SocialObject x))`

$SocRelationship(M) \wedge SocMarriage(sm) \wedge LegMarriage(lm) \wedge LegMarriage(lm') \wedge T(t) \wedge T(t')$	`(SocialRelationship m)`
	`(SocialMarriage sm)`
	`(LegalMarriage lm)`
	`(LegalMarriage lm1)`
	`(Time t)`
	`(Time t1)`

$PRE(M,t) \wedge PRE(M,t') \wedge PRE(sm,t) \wedge PRE(sm,t') \wedge PRE(lm,t) \wedge \neg PRE(lm,t') \wedge \neg PRE(lm',t) \wedge PRE(lm',t')$	`(present m t)`
	`(present m t1)`
	`(present sm t)`
	`(present sm t1)`
	`(present lm t)`
	`(not (present lm t1))`
	`(not (present lm1 t))`
	`(present lm1 t1)`

$lm \neq lm' \wedge CF(sm,M,t) \rightarrow CF(lm,M,t) \wedge CF(sm,M,t') \rightarrow CF(lm',M,t')$	`(not (= lm1 lm2))`
	`(if (classify sm m t) (classify lm m t))`
	`(if (classify sm m t1) (classify lm1 m t1))`

Table 2. CLIF translation of Case 5

The principles of these translations should be rather obvious, but let me emphasise the most salient ones:

- The DOLCE abbreviated predicates are unfolded, e.g., 'CF' is translated as 'classify';
- Apostrophes are converted into digits;
- Long conjunctions are split into self-standing conjuncts.

Now, the respective prompt is built by concatenating the English text and the CLIF translation - for the actual example, see the first part of the prompt in listing 3 below.

3.2. Test dataset

Since we do not yet have a sufficient amount of training data for the task at hand, we are not in a position to properly train or even fine-tune an LLM. Thus, we cannot simply send *any* natural language text with a couple of prompts to the model and expect an acceptable response. Instead, for the time being, we need to search for such natural language texts that are similar to the AO-examples. To achieve this purpose, i.e., to find such texts, I queried the LLM model, for each AO-example, with the following prompt:

```
In philosophical papers and books find 7 fragments that are most
similar to the text below. Answer with the actual quotes from
philosophy and not with their interpretations.
Text:
[AO-example's text goes in here]
```

Listing 1: 'Find similar texts' prompt template

The texts I got as responses from the LLM turned out to be:

- sufficiently similar to the AO-examples except for Case 2, where the LLM found texts like 'The only thing that is constant is change';
- of various degrees of dissimilarity within one batch;
- usually hallucinations, i.e., as a rule, they are not actual quotes from the philosophical texts they refer to, despite the explicit prompt.

(The latter observation is of minor importance to the current research as we are not after a search-and-quote assistant.)

Listing 2 shows the response I got for the query for Case 5 - all other responses can be found in the texts.json file in the GitHub repository.

```
1. Marriage, as we understand it in our society, is a mutual contract, usually between a man and a woman,
to live together as husband and wife.
– John Stuart Mill, The Subjection of Women
2. Marriage is a social institution as old as the human race itself.
It may be defined as a relation between a man and a woman which is recognized by custom or law
and involves certain rights and duties, both in the case of the parties entering the Union,
and in case of the children born of it.
– Bertrand Russell, Marriage and Morals
3. Marriage is a contract, one you enter into expecting to get your needs met during the good times and
    bad.
– Friedrich Nietzsche, Thus Spoke Zarathustra
4. Marriage is a contract, and a contract is a contract, and it is a contract between two people,
and it is a contract that is recognized by society, and it is a contract that can be broken,
and it is a contract that can be changed.
– Jean–Paul Sartre, Being and Nothingness
5. Marriage is a social contract between two individuals that unites their lives legally, economically,
and emotionally.
The contractual marriage agreement usually implies that the couple has legal obligations to each other
throughout their lives or until they decide to divorce.
– Immanuel Kant, The Metaphysics of Morals
6. Marriage is a contract between two persons which imposes mutual obligations and gives mutual rights.
It is a contract in which the parties undertake to share the rights and duties of domestic life.
– John Locke, Two Treatises of Government
7. Marriage is a contract for a permanent relationship, which is made in conformity with certain legal
    rules
and which may be dissolved only by judicial decision or under conditions provided for by law.
– Thomas Hobbes, Leviathan
```

Listing 2: Texts similar to Case 5

3.3. LLM queries

Finally, ontological prompts are concatenated with the appropriate natural language texts and sent as queries to the LLM. One of such queries is shown below:

```
English:
Marriage is a contract between two people that is present in most
social and cultural systems and it can change in major
(e. g. gender constraints) and minor (e.g. marriage breaking
procedures) aspects.
CLIF:
(if (SocialMarriage x) (Concept x))
(if (LegalMarriage x) (Concept x))
(if (SocialRelationship x) (SocialObject x))
(SocialRelationship m)
(SocialMarriage sm)
(LegalMarriage lm)
(LegalMarriage lm1)
(Time t)
(Time t1)
(present m t)
(present m t1)
(present sm t)
(present sm t1)
(present lm t)
(not (present lm t1))
(not (present lm1 t))
(present lm1 t1)
(not (= lm1 lm2))
(if (classify sm m t) (classify lm m t))
(if (classify sm m t1) (classify lm1 m t1))
English:
Marriage, as we understand it in our society, is a mutual
contract, usually between a man and a woman, to live together
as husband and wife.
CLIF:
```

Listing 3: Example of the ’formalisation prompt’

This text was directly sent to the LLM model.

4. Evaluation

Each response I got from the LLM - see the formalised_texts.json file in the repository - was evaluated with respect to the following features:

1. syntactic well-formedness, i.e., that whether a response is a well-formed CLIF theory;
2. internal consistency, i.e., that whether a response that is a well-formed CLIF theory is consistent;

Evaluation Criterion	Tested Count	Passed Count	Failed Count
syntactic well-formedness	42	38	*4*
internal consistency	38	37	1
DOLCE consistency	37	37	0
semantic adequacy	37	23	14

Table 3. Formalisations' evaluation summary

3. DOLCE consistency, i.e., that whether a response that is a well-formed CLIF theory is consistent with the DOLCE ontology;
4. semantic adequacy with respect to the DOLCE ontology, i.e., that whether the CLIF theory conveys a similar meaning as the original English text in the context of the DOLCE ontology.

The criteria 1–3 are purely objective and can be automated.[5] Criterion 4 was not automated, mainly because it requires human insight into the philosophical assumptions of DOLCE. Also, as of now, it is helplessly vague, i.e., I am not in a position to operationally specify in this paper what it means that a CLIF theory conveys a similar meaning to an English text in the context of a given ontology.

The table 3 summarises the results of this evaluation by giving the relevant stats - the actual results are stored in the outputs folder in the GitHub repository. Let me focus on the outliers here, i.e., on those cases where the process did not deliver the expected results. First, there are four English texts for which the process did not produce well-formed CLIF texts:

```
1. The rose is red in the morning, but as the day wears on, the
color fades. By evening, the rose is a dull, lifeless brown.

2. The cherry blossom is pink in the spring. As the seasons
change, so too does its hue. In the fall, the cherry blossom is
a muted brown.

3. We must be willing to let go of the life we planned so as to
have the life that is waiting for us.

4. Life is a series of natural and spontaneous changes. Don't
resist them; that only creates sorrow. Let reality be reality.
Let things flow naturally forward in whatever way they like.
```

Listing 4: Texts whose formalisations fail the syntactical criterion

Note that the last two texts are not similar to any of the AO-examples - indeed, when prompted by them, the LLM responded that it cannot translate them into CLIF. So, in a sense, they are not LLM failures but just explicitly indicate the limits of the approach I outline here.

As for the one well-formed CLIF theory that is inconsistent, it was produced as the translation of the text shown in listing 5.

[5] To check the well-formedness I adapted the code of Macleod parser. The consistency results were generated by the Vampire prover using the TPTP translations of the CLIF texts from the LLM. The DOLCE ontology in the TPTP format was generated in the same way as described in [13].

AO-example	Tested Count	Passed Count	Failed Count
Case 1: Composition/constitution	6	4	2
Case 2: Roles	5	0	5
Case 3.1: Color change	5	5	0
Case 3.2: Speed change	7	7	0
Case 4: Event change	7	7	0
Case 5: Concept evolution	7	0	7

Table 4. Semantic adequacy results breakdown by AO-example

```
The snake which cannot cast its skin has to die.
As well the minds which are prevented from changing their
opinions; they cease to be mind.
```

Listing 5: Text formalised as an inconsistent theory

The resulting inconsistent CLIF theory is given in listing 6.

```
(if (Organism x) (PhysicalObject x))
(if (MentalProcess x) (Process x))
(if (and(Organism x) (hasProperty x y t) (hasProperty x1 y t1))
(= x x1))
(if (and(MentalProcess x) (hasProperty x y t)
(hasProperty x1 y t1))
(= x x1))
(Organism Snake)
(MentalProcess Mind)
(Property SkinCasting)
(Property OpinionChanging)
(not (Property SkinCasting))
(not (Property OpinionChanging))
(Time t1)
(Time t2)
(< t1 t2)
(hasProperty Snake SkinCasting t1)
(hasProperty Mind OpinionChanging t1)
(forall (x) (not (hasProperty x SkinCasting t2)))
(forall (x) (not (hasProperty x OpinionChanging t2)))
(not (Organism Snake t2))
(not (MentalProcess Mind t2))
```

Listing 6: Inconsistent formalisation

Again, the English text is dissimilar to all AO-examples.

As for the semantic adequacy, let me first note that different AO-examples produce radically different results - see the results' breakdown in table 4. The LLM handled the texts similar to Cases 3.1, 3.2, 4, and, to a certain extent, Case 1 quite well, but failed for Cases 2 and 5.

Consider one of the failing translations for the Case 5 type:

```
Marriage is a contract for a permanent relationship , which is made
in conformity with certain legal rules and which may be dissolved
only by judicial decision or under conditions
provided for by law .
```

Listing 7: Example of a text that was not adequately formalised

The LLM translated this text as shown by listing 8:

```
(if (PermanentRelationship x) (Concept x))
(if (LegalRules x) (Concept x))
(if (JudicialDecision x) (Concept x))
(if (LegalConditions x) (Concept x))
(PermanentRelationship pr)
(LegalRules lr)
(JudicialDecision jd)
(LegalConditions lc)
(LegalMarriage lm)
(Time t)
(Time t1)
(present pr t)
(present pr t1)
(present lr t)
(present lr t1)
(present jd t)
(not (present jd t1))
(present lc t)
(not (present lc t1))
(if (classify lm pr t) (classify lm lr t))
(if (classify lm pr t1) (classify lm jd t1))
(if (classify lm pr t1) (classify lm lc t1))
```

Listing 8: Inadequate formalisation example

Although this translation has captured some salient ontological aspects of the situation the English text describes, it has missed quite a few others. So, for example, it seems that it correctly classified such objects as legal rules as concepts, but it missed that the marriage contract may be dissolved at a certain time under specific conditions.

One may wonder why the formalisations of the texts similar to Cases 3.1, 3.2, and 4 were so much better than the formalisations of other texts. A seemingly reasonable explanation of this fact can be derived from the following two comparisons of textual similarities:

1. comparison between the original Case 5 and the texts found by the LLM - see listing 2 above;
2. comparison between the original Case 3.1 and the texts found by the LLM - see listing 9 below.[6]

[6]To save space, I left out from listing 9 all references to the authors and works quoted.

AO-example	Average similarity for found texts
Case 1	0.913
Case 2	0.725
Case 3.1	0.963
Case 3.2	0.924
Case 4	0.964
Case 5	0.804

Table 5. Text average similarities by AO-examples

Intuitively, the similarities in 1 are far more remote than the ones found in 2. Alternatively, one can use one of many NLP methods of text comparison and get less subjective results - table 5 shows the average measurements of the text similarities between the AO-examples and the texts found by the LLM, which were spawned by a word2vec model of English, i.e., by the en_core_web_lg model of the spacy library (`https://spacy.io`).[7]

So one may venture to claim that the quality of the LLM translation of an English text to CLIF in the context of a formal ontology depends on the similarity of this text to the text for which the CLIF formalisation was crafted by a human. However, since the amount of data I have at my disposal is statistically insignificant, this comparison may be seen as insufficient evidence for this hypothesis.

```
1. The rose is red in the morning, but as the day wears on, the
   color fades. By evening, the rose is a dull, lifeless brown.
2. The cherry blossom is pink in the spring. As the seasons
   change, so too does its hue. In the fall, the cherry blossom
   is a muted brown.
3. A tulip is vibrant in the summer. As the months roll by, its
   color alters. In the autumn, the tulip is a faded brown.
4. The daisy is white in the summer. As the days pass, its color
   shifts. In the fall, the daisy is a dull brown.
5. The sunflower is yellow in the summer. As the weeks pass, its
   color transforms. In the fall, the sunflower is a deep brown.
6. The poppy is bright in the summer. As the season changes, so
   too does its color. In the fall, the poppy is a dark brown.
7. The lily is pure in the summer. As the days shorten, its color
    changes. In the fall, the lily is a somber brown.
```

Listing 9: Texts similar to Case 3.1

In sum, 23 out of 42 texts, i.e., approx. 55% of all texts, were adequately formalised as logical theories based on and consistent with the DOLCE ontology. In my view, this finding justifies the claim that a large language model can understand some, but obviously not all, of DOLCE.

[7] All similarity measurements can be found in the texts_similarities.json file in the repository. Needless to say, other embeddings and other NLP methods may, in general, yield different results - for a recent survey of methods to compare short texts, see: [15].

5. Conclusions

As may have been expected, a large language model may not be, out of the box, capable of producing logical artefacts at a satisfactory level, even if it is capable of generating XML documents or SPARQL queries. Nonetheless, the research experiment presented in this paper indicates that one can prompt it to produce ontologically adequate descriptions of certain situations, provided that one possesses a human-made ontological description of a sufficiently similar situation.

One may also expect that fine-tuning a large language model may increase its capability to produce ontologically relevant formalisations of natural language descriptions. The main obstacle here would be the scarcity of the training data, as we do not yet have a sufficient amount of such translations. The method used in [2] to extend the FOLIO dataset can be adapted for this purpose. Another route may lead through the intermediate step of a controlled natural language suitable for such languages as CLIF, as recommended in [1].

References

[1] Lehmann J, Gattogi P, Bhandiwad D, Ferré S, Vahdati S. Language Models as Controlled Natural Language Semantic Parsers for Knowledge Graph Question Answering. In: ECAI 2023 - 26th European Conference on Artificial Intelligence. vol. 372 of Frontiers in Artificial Intelligence and Applications. Krakow (Cracovie), Poland: IOS Press; 2023. p. 1348-56.

[2] Yang Y, Xiong S, Payani A, Shareghi E, Fekri F. Harnessing the Power of Large Language Models for Natural Language to First-Order Logic Translation; 2023. https://arxiv.org/abs/2305.15541.

[3] Han S, Schoelkopf H, Zhao Y, Qi Z, Riddell M, Benson L, et al.. FOLIO: Natural Language Reasoning with First-Order Logic; 2022. https://arxiv.org/abs/2209.00840.

[4] Xu F, Lin Q, Han J, Zhao T, Liu J, Cambria E. Are Large Language Models Really Good Logical Reasoners? A Comprehensive Evaluation and Beyond; 2023. https://arxiv.org/abs/2306.09841.

[5] Chen M, Ma Y, Song K, Cao Y, Zhang Y, Li D. Learning To Teach Large Language Models Logical Reasoning; 2023. https://arxiv.org/abs/2310.09158.

[6] Su Y, Fu X, Liu M, Guo Z. Are LLMs Rigorous Logical Reasoner? Empowering Natural Language Proof Generation with Contrastive Stepwise Decoding; 2023. https://arxiv.org/abs/2311.06736.

[7] Neuhaus F. Ontologies in the era of large language models – a perspective. Applied Ontology. 2023;18(4):399–407.

[8] Mateiu P, Groza A. Ontology engineering with Large Language Models; 2023. https://arxiv.org/abs/2307.16699.

[9] OpenAI. GPT-4 Technical Report; 2023. https://arxiv.org/abs/2303.08774.

[10] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. Advances in neural information processing systems. 2020;33:1877-901.

[11] Borgo S, Galton A, Kutz O, Borgo S, Galton A, Kutz O. Foundational Ontologies in Action. Applied Ontology. 2022;17(1):1–16.

[12] Borgo S, Ferrario R, Gangemi A, Guarino N, Masolo C, Porello D, et al. DOLCE: A descriptive ontology for linguistic and cognitive engineering. Applied Ontology. 2022;17(1):45-69.

[13] Garbacz P. Foundational Ontologies Meet Automatic Theorem Provers. In: Sales TP, Hedblom MM, Tan H, Righetti G, Kutz O, Glauer M, et al., editors. Proceedings of the Joint Ontology Workshops 2022 Episode VIII: The Svear Sommar of Ontology, Jönköping, Sweden, August 15-19, 2022. vol. 3249 of CEUR Workshop Proceedings. CEUR-WS.org; 2022. .

[14] Information technology — Common Logic (CL) — A framework for a family of logic-based languages. International Organization for Standardization; 2018.

[15] Han M, Zhang X, Yuan X, Jiang J, Yun W, Gao C. A survey on the techniques, applications, and performance of short text semantic similarity. Concurrency and Computation: Practice and Experience. 2021;33(5):e5971.