

Article

Not peer-reviewed version

---

# Ontology-Driven Semantic Alignment: Assessing the Reasoning Capabilities of Large Language Models in Geospatial Contexts

---

[Fabíola Andrade Souza](#) \* and [Silvana Philippi Camboim](#)

Posted Date: 22 December 2025

doi: 10.20944/preprints202512.1800.v1

Keywords: semantic interoperability; Large Language Models (LLM); neuro-symbolic AI; geospatial ontologies; Spatial Data Infrastructure (SDI); OpenStreetMap



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Ontology-Driven Semantic Alignment: Assessing the Reasoning Capabilities of Large Language Models in Geospatial Contexts

Fabiola Andrade Souza <sup>1,2,\*</sup> and Silvana Philippi Camboim <sup>2</sup>

<sup>1</sup> Universidade Federal da Bahia, Brazil

<sup>2</sup> Universidade Federal do Paraná, Brazil

\* Correspondence: [fabiola.andrade@ufba.br](mailto:fabiola.andrade@ufba.br)

## Abstract

Semantic interoperability remains a critical challenge in Spatial Data Infrastructures (SDIs), particularly when aligning authoritative taxonomies with collaborative folksonomies. Recent advances in Large Language Models (LLMs) offer new avenues for automated semantic interpretation, yet these 'sub-symbolic' approaches often lack the logical rigor required for structured geospatial data. This paper evaluates the capability of LLMs – specifically distinguishing between traditional architectures and emerging Large Reasoning Models (LRMs) – to perform semantic alignment between the Brazilian national cartographic standard (EDGV) and OpenStreetMap (OSM). Using a formal ontology as a prompting scaffold, we tested seven model versions (including ChatGPT 5.0, DeepSeek R1, and Gemini 2.5) on their ability to identify semantic equivalents and generate valid ontological mappings. Results indicate that while traditional LLMs struggle with hierarchical structures, reasoning-oriented models demonstrate significantly improved capacity for complex inference, correctly identifying many-to-one (n:1) relationships across linguistic barriers. However, all models exhibited limitations in generating syntactically valid OWL code, revealing a gap between semantic comprehension and formal structuring. We conclude that a neuro-symbolic approach, using ontologies to ground AI reasoning, provides a viable pathway for semi-automated interoperability, although future work must address the lack of explicit spatial reasoning in current architectures.

**Keywords:** semantic interoperability; Large Language Models (LLM); neuro-symbolic AI; geospatial ontologies; Spatial Data Infrastructure (SDI); OpenStreetMap

## 1. Introduction

The representation of geospatial phenomena has long relied on conceptual models that formalize how entities, attributes, and relationships should be interpreted within a given domain. In Spatial Data Infrastructures (SDIs), where data from multiple institutions and scales converge, the lack of shared conceptualizations frequently leads to semantic heterogeneity and obstructs interoperability. This challenge becomes particularly significant when integrating authoritative geospatial datasets – such as Brazil's official topographic specification (EDGV) – with collaboratively produced structures, such as OpenStreetMap (OSM).

Semantic discrepancies between these models arise from linguistic conventions, cultural abstraction, and domain-specific categorization. Human cognition adds another layer of complexity: conceptualizations reflect hierarchical categorization, prototype structures, and culturally embedded meanings [1–4]. As a result, aligning heterogeneous schemas requires more than lexical matching; it demands understanding hierarchical structures, domain constraints, and context-dependent meanings. Ontologies, therefore, play a central role in structuring conceptual knowledge and mitigating semantic inconsistencies across geospatial datasets [5,6].

Advances in Natural Language Processing (NLP) and Large Language Models (LLMs) have recently created new opportunities for supporting semantic interoperability. Because map legends, data dictionaries, and model definitions are primarily expressed in natural language, LLMs can help infer semantic correspondences across schemas, including those expressed in different languages or at different abstraction levels. Preliminary studies indicate that LLMs can recognize similarities between geospatial concepts when these are presented as structured text or formal ontologies [7–9]. However, the extent to which LLMs – and more recent reasoning-oriented variants (Large Reasoning Models, LRMs) – can understand and reason over formalized geospatial ontologies remains insufficiently explored.

Existing research has extensively employed ontology-driven methods for semantic alignment using traditional similarity metrics [10–13]. While recent studies have evaluated the quality of ontological abstractions [14], there is still a gap in assessing how LLMs handle hierarchical and relational reasoning within these structured representations. Simultaneously, current literature highlights critical limitations in LLM reasoning, including weaknesses in logical consistency and spatial understanding [15–18]. These findings raise important questions about the suitability of current geospatial alignment architectures.

To address this gap, this study evaluates whether LLMs, both traditional and reasoning-oriented, can interpret an OWL ontology derived from the EDGV schema and establish semantic correspondences with OSM tags. Specifically, this study aims to: The analysis examines the models' ability to: (i) comprehend hierarchical structures; (ii) generate semantically coherent alignments across multilingual schemas; and (iii) reproduce the ontology in valid OWL syntax. By assessing seven versions of three model families (ChatGPT, DeepSeek, and Gemini), this paper provides an empirical evaluation of their capabilities and limitations for supporting semantic interoperability in geospatial data integration.

## 2. Related Work and Conceptual Background

### 2.1. Authoritative Geospatial Models: EDGV

Brazil's 'Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais' (EDGV) is a prescriptive, hierarchical, and strongly typed national specification designed to promote semantic consistency and interoperability across federal geospatial producers [19]. Its conceptual model is organized into domains, categories, and classes, each defined by controlled vocabularies and attribute domains. This structure reflects a top-down governance model aligned with symbolic data-modelling traditions, where each feature class is supported by explicit constraints that guide both data production and validation. For example, a "hotel" is not stored as a free-text label but as an instance of the class Commerce and/or Services Building, associated with a specific, predefined attribute value indicating its subtype [6]. Such granularity enables rigorous, automated quality assurance and supports national-scale interoperability, but also produces rigid conceptual boundaries that complicate integration with more flexible datasets.

### 2.2. OpenStreetMap: A Collaborative Folksonomy

In contrast to EDGV's rigidity, OpenStreetMap (OSM) constitutes a decentralized, volunteer-driven mapping initiative governed by community consensus rather than formal standards. Its data model is intentionally lightweight, comprising nodes, ways, and relations whose semantics are expressed entirely through an open-ended system of tags. These key=value pairs allow contributors to describe features with remarkable flexibility, from general tags such as building=yes to highly specific attributes that capture local variations [20–22]. The system supports unlimited multilingual tag values, and its semantics evolve dynamically through use and documentation on the OSM Wiki. As a result, OSM often offers more up-to-date, fine-grained, and context-sensitive information than official maps, particularly in urban centers. However, this flexibility comes at the cost of structural

and semantic uniformity, generating inconsistencies and ambiguities that pose substantial challenges for alignment with authoritative standards such as EDGV [6].

### 2.3. The Semantic Alignment Challenge

Aligning EDGV and OSM involves more than translating terms or matching database structures. It requires reconciling two fundamentally different conceptual models driven by distinct philosophies of data creation. EDGV embodies a prescriptive, Portuguese-language, hierarchical worldview designed for institutional consistency, whereas OSM reflects an emergent, largely English-dominant folksonomy grounded in collaborative, bottom-up practices. This divergence introduces linguistic, structural, and conceptual mismatches. A single EDGV class may correspond to multiple highly specific OSM tags. Conversely, a detailed OSM feature may need to be decomposed and reinterpreted to fit into EDGV's rigid schema. More deeply, the two models categorize and describe the world through different epistemological lenses: EDGV defines fixed categories sanctioned by national agencies. At the same time, OSM represents how contributors perceive and describe features in situ. Achieving meaningful interoperability, therefore, requires navigating not only naming differences but also conflicting modelling assumptions about how geographic phenomena are structured and understood.

This challenge has significant practical implications for Brazilian Spatial Data Infrastructures. OSM can complement gaps in official mapping, particularly in fast-changing areas, but only if robust semantic interoperability mechanisms bridge the conceptual chasm between these two models.

### 2.4. AI Paradigms and Ontological Reasoning

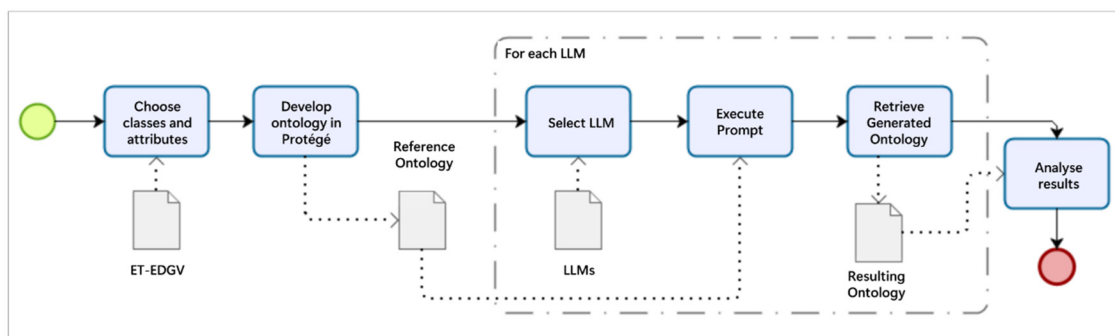
Artificial Intelligence provides several pathways for supporting semantic alignment, each grounded in distinct theoretical traditions. Historically, AI has evolved through symbolic, connectionist, and, more recently, neuro-symbolic paradigms [23,24]. Symbolic methods rely on formal representations, logical reasoning, and explicit rules, making them suitable for modelling geospatial concepts whose structure must be verifiable and auditable [23]. However, symbolic systems depend heavily on manually encoded knowledge and scale poorly in the face of linguistic variability.

The rise of connectionist (sub-symbolic) approaches, particularly neural networks and Transformer-based language models, shifted the focus to learning patterns from large corpora [25–27]. Large Language Models (LLMs) excel at capturing linguistic regularities and generating plausible text but often struggle with hierarchical reasoning, logical consistency, and systematic generalization [16,18,28]. These weaknesses are especially pronounced in tasks involving structured knowledge, such as ontologies.

Ontologies became central during the Semantic Web era as a means to formalize domain knowledge through classes, properties, and axioms [29,30]. While they enable explicit reasoning and consistency checking, their interpretative scope is limited to what is explicitly modelled. Geospatial data, typically expressed through hybrid notations that combine text, geometry types, and hierarchical relationships, presents challenges for both purely statistical and purely symbolic systems [15,31,32]. This has renewed interest in neuro-symbolic approaches that combine the linguistic breadth of LLMs with the logical rigor of formal ontologies [23].

## 3. Methodology

This study investigates the capacity of Large Language Models (LLMs) to reason about formal geospatial ontologies. Specifically, we evaluate whether these tools can identify semantically equivalent concepts across two structurally divergent schemas: the prescriptive Brazilian EDGV and the community-driven OSM folksonomy. The experimental workflow (Figure 1) comprises three stages: (i) formalizing the EDGV schema into an ontology; (ii) prompting diverse LLMs to align this ontology with OSM tags; and (iii) evaluating the semantic and structural quality of the outputs.



**Figure 1.** Flowchart of the methodology.

### 3.1. Ontology Construction

The ‘Buildings’ category of the EDGV specification was selected as the reference domain, comprising 14 georeferenced classes, attributes, and domains (see Table 1). To enable symbolic reasoning, this schema was modelled in OWL using Protégè (version 5.6.4). Crucially, the ontology relies solely on descriptive semantics (taxonomies and properties), excluding geometric primitives. This design choice mitigates known limitations of current LLMs in handling spatial topology [15,17,32] and focuses the evaluation on semantic inference.

**Table 1.** Classes, attributes, and domains from the Buildings Category of EDGV used.

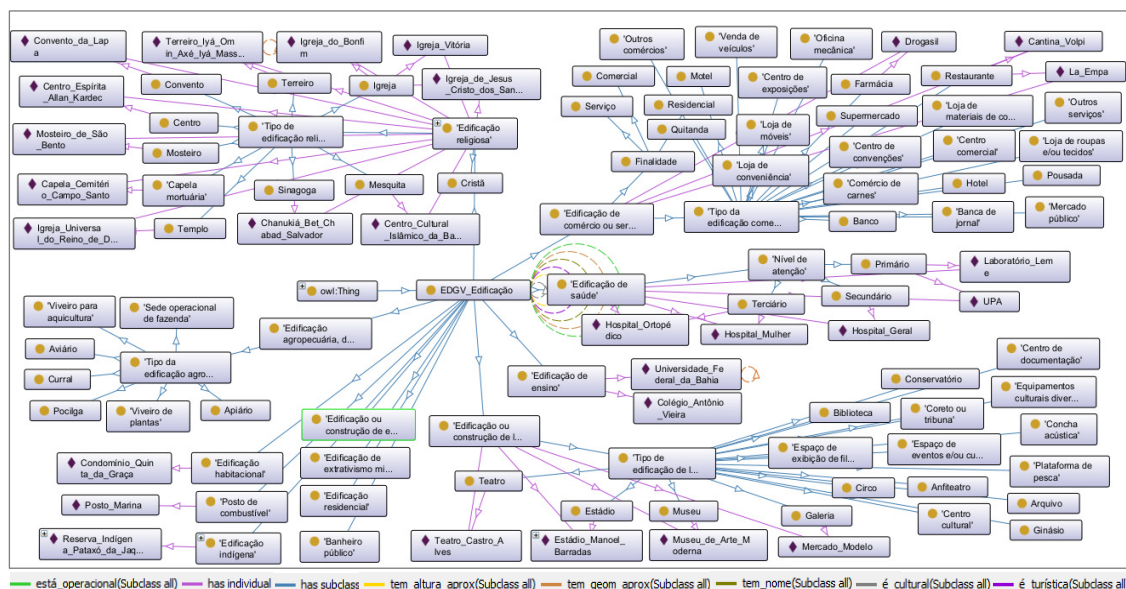
Class	Attribute	Domain
Building	Name	-
	Approximate geometry	-
	Operational	Yes / No / Unknown
	Approximate height	-
	Touristic	Yes / No / Unknown
	Cultural	Yes / No / Unknown
Fuel station	-	-
Public toilets	-	-
Educational Building	-	-
Farming, plant extraction and/or fishing Building	Building type	Apiary
		Aviary
		Barn
		Pigsty
		Farm operational headquarters
		Plant nursery
		Aquaculture nursery
Commerce and/or Services Building	Building type	Commercial
		Residential
		Services
		Newsstand
		Bank
Commerce and/or Services Building	Building type	Shopping center
		Convention center
		Exhibition center
		Butcher shop
		Pharmacy
		Hotel
		Convenience store
Commerce and/or Services Building	Building type	Building materials and/or hardware store
		Furniture store
		Clothing and/or fabric store

		Public marketplace Motel Car Repair Other businesses Other services Inn Greengrocer Restaurant Supermarket Dealership
Mineral extraction Building	-	-
Healthcare Building	Level of care	Primary Secondary Tertiary
Housing Construction	-	-
Indigenous Building	Collective	-
	Isolated	-
Residential Building	-	-
Building or Construction of a phenomenon measurement station	-	-
		Amphitheater Public Records Library Cultural Center Documentation center Circus Acoustic concert hall Conservatory Bandstand Various cultural facilities Event and/or cultural space Film screening space Stadium Gallery Gymnasium Museum Fishing platform Theater
Building or Construction of leisure	Building type	
	Christian	-
	Teaching	-
	Religion type	-
Religious Building	Building type	Mortuary chapel Center Convent Church Mosque Monastery Synagogue Temple Afro-Brazilian religious temple ('Terreiro')

Source: Adapted from [19] – original in Brazilian Portuguese.

Modelling adaptations were required to bridge the gap between the rigid, scale-dependent EDGV structure using the OMT-G data model [33] and the flexibility required for alignment. 'EDGV\_Building' was defined as the root class. Due to reasoning limitations regarding complex property constraints, attributes and domains were modelled as nested subclasses (e.g., 'Healthcare

Building' à 'Level of Care' à 'Primary'). These processes resulted in a hierarchy of 87 class elements. Additionally, an 'OSM\_Tags' superclass was established to house the alignment targets. Instances representing real-world locations in Salvador, Brazil, were populated to test instance-level reasoning. The resulting EDGV ontology serves as the ground truth for the alignment experiments and is available in the project repository (see Data Availability Statement). To inspect the resulting taxonomy and class relationships, the ontology was visualized using the OntoGraf plugin for Protégè, as illustrated in Figure 2.



**Figure 2.** Visualization of the reference EDGV ontology structure generated using the OntoGraf plugin.

Unlike EDGV, OpenStreetMap lacks a formal schema. Consequently, no input ontology was created for OSM; instead, the LLMs were tasked with retrieving concepts directly from their internal knowledge of the OSM Wiki folksonomy [22], simulating a dynamic integration scenario.

### 3.2. Language Models and Dialogue Prompt

The NLP tools selected for the analysis were ChatGPT<sup>1</sup>, DeepSeek<sup>2</sup>, and Gemini<sup>3</sup>. All three tools are proprietary; however, DeepSeek provides openly accessible free versions, while Gemini and ChatGPT offer both free and paid tiers.

Different versions of each tool were employed, including ChatGPT 4o, o1.preview, and 5.0; DeepSeek V3 and R1; and Gemini 2.0 Flash Thinking Experimental and 2.5 Pro. While strictly speaking, all evaluated models are Large Language Models (LLMs) based on Transformer architectures, we distinguish between General-Purpose LLMs (e.g., ChatGPT 4o, DeepSeek V3) and Large Reasoning Models (LRMs) (e.g., ChatGPT 5.0, DeepSeek R1, Gemini 2.5 Pro). The latter are specifically optimized via reinforcement learning to simulate methodical cognitive processes (Chain-of-Thought) before generating an output.

The selection of these tools was based on several criteria. ChatGPT was included due to its strong performance in earlier studies on semantic reasoning for geospatial ontologies [8,9]. DeepSeek and Gemini were selected primarily because they represent distinct model architectures relevant for comparison.

For the evaluation, all seven model versions received the same task prompt and the same ontology code. The prompt instructed the models to identify the semantic associations between subclasses of the EDGV\_Edificações class and the corresponding OpenStreetMap (OSM) tags, and to generate new subclasses within the OSM\_Tags class using appropriate OWL notation. This prompt (see Listing 1) was used verbatim in all experiments to ensure full reproducibility.

**Listing 1.** Prompt used in all experiments.

Considering that the instances in the ontology belong to classes and subclasses of the EDGV\_Edificações class, for example, the UPA instance is of the Primário and Edificação de Saúde types.

Considering also that all the classes and subclasses presented in the ontology inherit from EDGV\_Edificações and can have OSM tags semantically associated with them (according to existing tags on [https://wiki.openstreetmap.org/wiki/Map\\_features](https://wiki.openstreetmap.org/wiki/Map_features)).

Please make the semantic association of all the subclasses of EDGV\_Edificações in this ontology with the OSM tags on the website, if compatible.

As a result, create new subclasses for the OSM\_Tags class with the name of the OSM tags that should be used to complement the ontology, indicating their semantic association with the corresponding subclass(es) of EDGV\_Edificações. Use the appropriate OWL notation.

### 3.3. Evaluation Protocol

Generated ontologies were analyzed structurally using the Protégé software and visualized using the OntoGraph tool, with findings detailed in section 4 and the discussion in section 5. LLM's performance against the dialogue prompt (Listing 1) was evaluated using a set of metrics covering multiple aspects:

- **Completeness in Semantic Alignment:** number of classes accurately associated in relation to the total of 87 classes, attributes, and domains of the EDGV represented in the input ontology, considering the readability of the EDGV ontology and the OSM wiki. This analysis assessed geospatial data quality, specifically the completeness dimension, drawing on ISO standards [34]. The criteria for determining a commission were whether or not the association was considered appropriate to the ontology class, and for omission, the failure to make an association to the ontology class. The metric for determining the appropriate association was based on the authors' observation and the conceptual definition of the elements in their original schemas.
- **Syntactic Conformity:** the ability to generate valid OWL code while maintaining structural integrity, for example, preserving the existing class hierarchy, consistently creating new classes representing the OSM tags as subclasses of OSM\_Tags, maintenance of the original object properties and instances, and use of 'EquivalentTo' notation as an element to semantic association.
- **Complex reasoning:** test of LLM's ability to infer domain rules, by making associations at different hierarchical levels (e.g. all religious buildings correspond to 'amenity=place\_of\_worship'). Additionally, the ability to classify relationships into multiple associations (1:1, 1:n, and n:1) and identify the classes involved in many-to-one relationships.

This process evaluates model evolution (between versions of the same LLM) and variation (between different LLMs), as well as their typical failure modes. The goal was to isolate the impact of incremental training and each LLM's architecture on the ability to understand the ontology and

generate semantic alignment, as measured by the applied metrics. Ultimately seeking to understand the advantages and limitations of integrating symbolic AI (ontology) with sub-symbolic AI (direct text dialogue), in a neuro-symbolic approach.

## 4. Results

The resulting ontologies for each model are available in the supplementary material (see Data Availability Statement). This section details the structural and semantic outcomes categorized by model architecture.

### 4.1. OpenAI (ChatGPT)

The traditional model (ChatGPT 4o) produced minimal alignment, identifying only four OSM tags (e.g., healthcare=hospital, tourism=hotel). Structurally, it failed to preserve the input ontology's hierarchy, properties, or instances, resulting in a flat list of simple 1:1 associations using the 'EquivalentTo' notation (Figure 3).

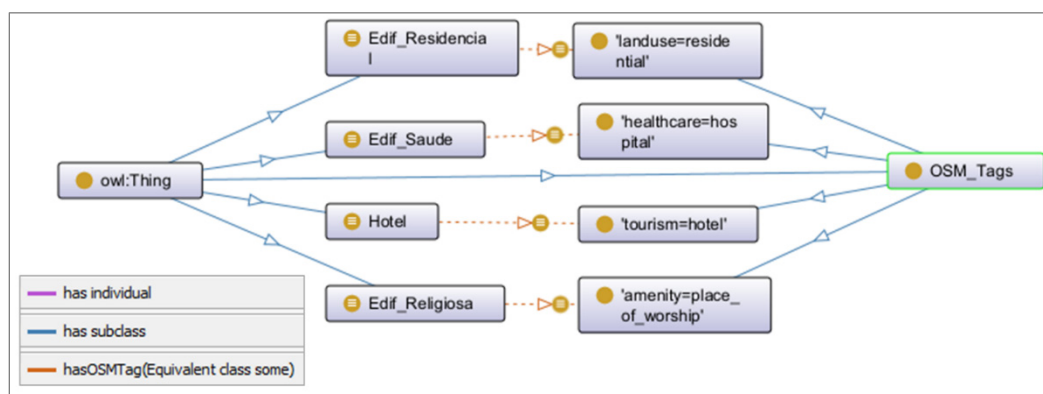


Figure 3. Ontological alignment generated by ChatGPT 4o (Traditional LLM).

A clear evolution was observed in the reasoning models. ChatGPT o1-preview increased the alignment to 14 associations but still discarded the original hierarchy and annotations (Figure 4). Conversely, the most advanced version (ChatGPT 5.0) achieved 31 associations and successfully maintained the original class hierarchy, object properties, and instances (Figure 5). Notably, it began to demonstrate complex reasoning, identifying 1:n relationships (e.g., relating the 'Hotel' domain to both amenity=hotel and building=hotel), although it duplicated the structural tree in the output.

### 4.2. DeepSeek

DeepSeek V3 (Traditional) struggled with OWL syntax, generating only 11 associations and inverting the hierarchy, placing the original EDGV classes as subclasses of OSM tags (Figure 6). Furthermore, it associated concepts only at the key level (e.g., educational building linked simply to amenity), ignoring specific values.

The reasoning model (DeepSeek R1) significantly improved semantic recall, identifying 34 associations, including complex n:1 relationships (e.g., grouping multiple healthcare levels under amenity=hospital), and consistently applied the 'EquivalentTo' axioms directly to class names (Figure 7). However, structural integrity remained poor; the model failed to preserve the original hierarchy and properties.

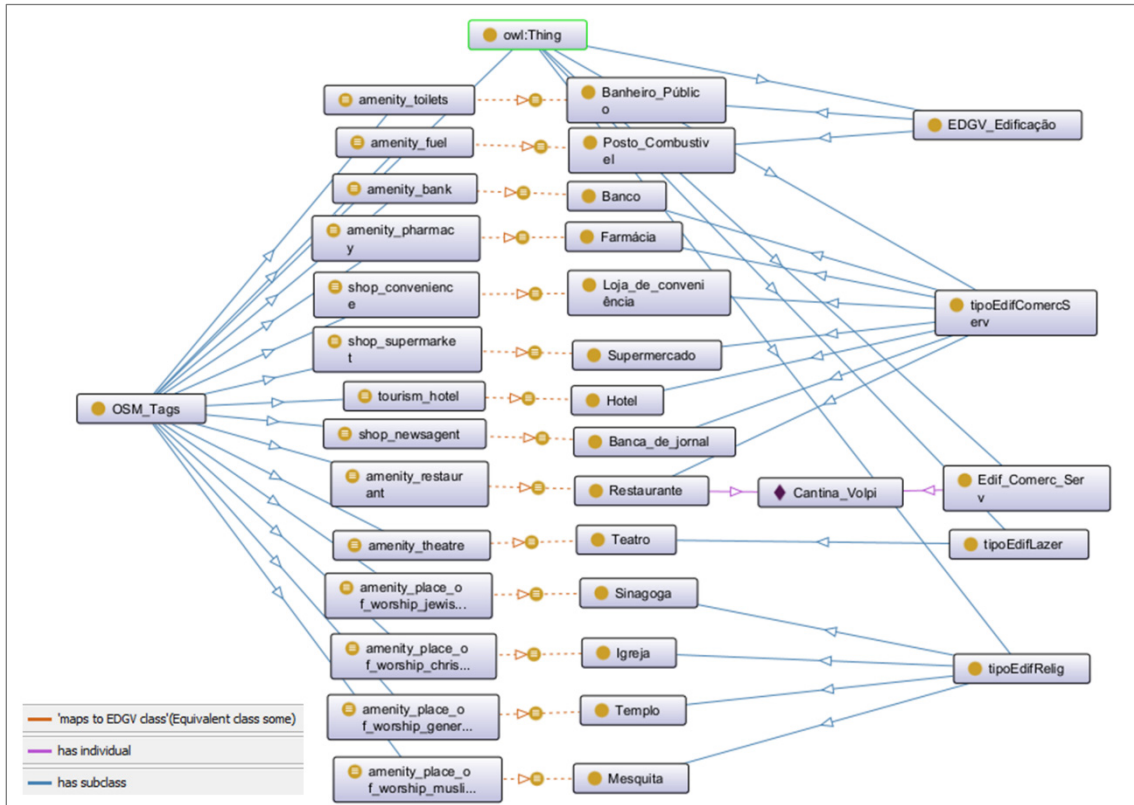


Figure 4. Ontological alignment generated by ChatGPT o1-preview (Reasoning Model).

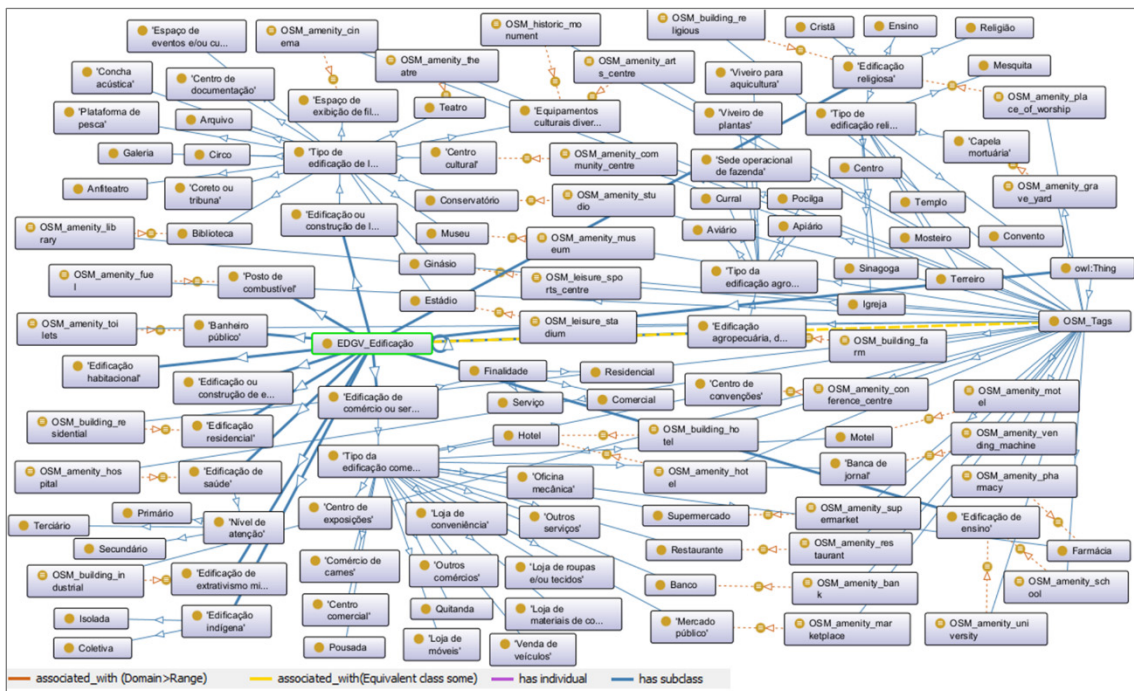


Figure 5. Ontological alignment generated by ChatGPT 5.0 (Reasoning Model).

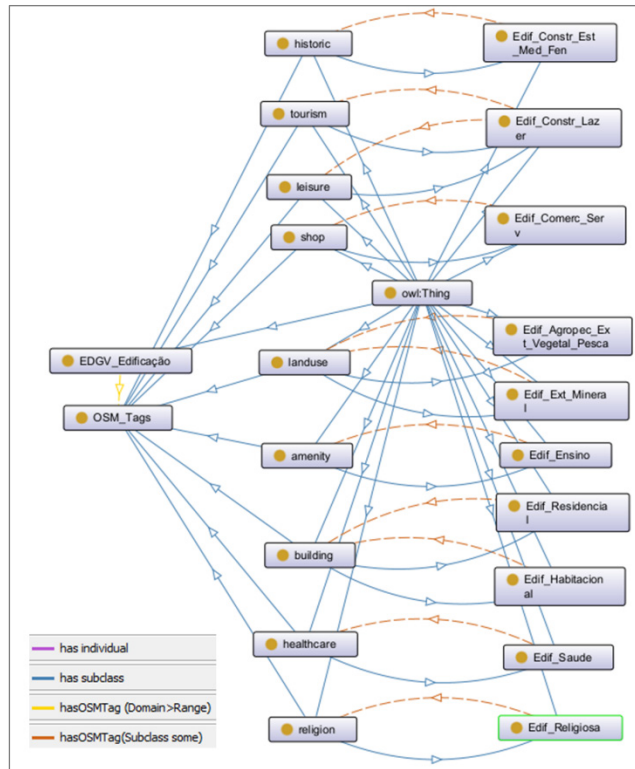


Figure 6. Ontological alignment generated by DeepSeek V3 (Traditional LLM).

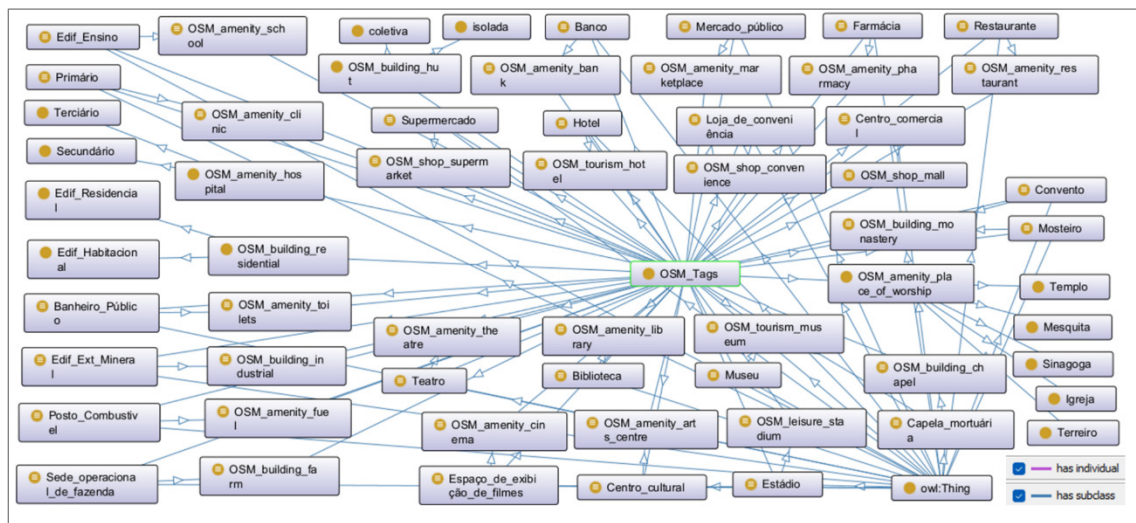


Figure 7. Ontological alignment generated by DeepSeek R1 (Reasoning Model).

### 4.3. Google (Gemini)

Gemini models achieved the highest recall, identifying 62 semantic associations in both versions. The preview model (2.0 Flash Thinking) exhibited structural flaws similar to DeepSeek, inverting the hierarchy and merging class names (e.g., OSM\_Tags\_amenity\_hospital) (Figure 8).



significantly in interpreting these structures, a gap remains between semantic recognition and logical syntactic generation.

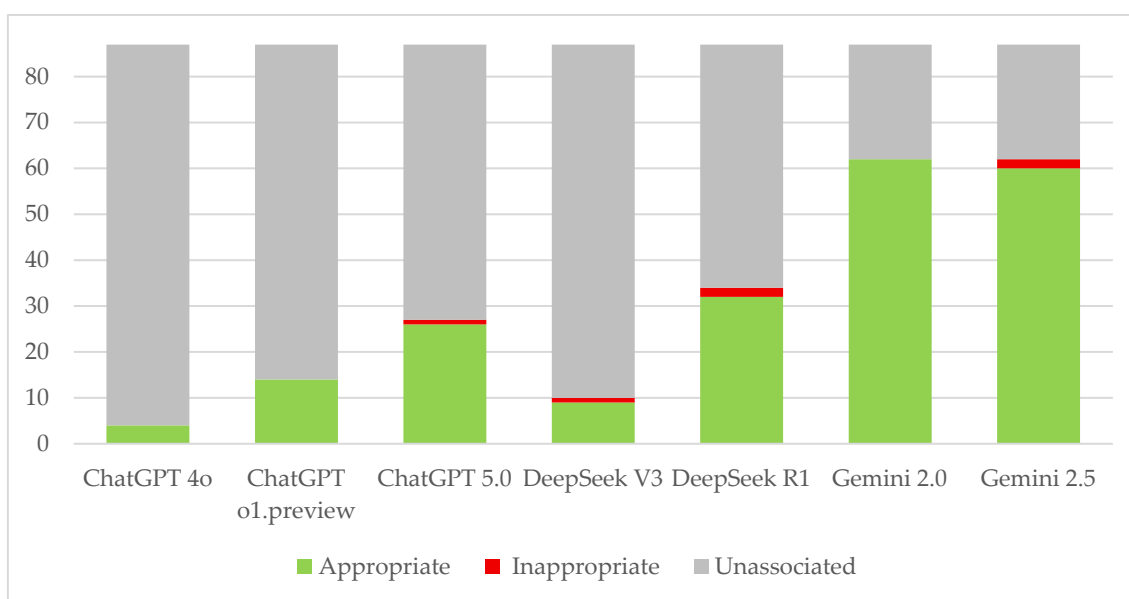
### 5.1. Completeness in Semantic Alignment

As detailed in Table 2, reasoning models (LRMs) consistently outperformed traditional architectures in semantic recall. Gemini (versions 2.0 and 2.5) achieved the highest completeness (~70% of classes associated), a substantial improvement over ChatGPT 4o's conservative performance (~4.6%). Crucially, the models successfully navigated the cross-lingual barrier (Portuguese-English) without explicit translation steps, corroborating the findings of [7]. While Table 2 summarizes the quantitative results, the complete list of semantic associations derived by each model – including the specific OSM tags and logical axioms employed – is detailed in the Supplementary Material.

**Table 2.** Number of EDGV ontology classes semantically associated with OSM tags.

Metric/ Category	ChatGPT			DeepSeek		Gemini	
	4o	o1.preview	5.0	V3	R1	2.0	2.5
<b>Unassociated classes (Omission)</b>	83 (95.4%)	73 (83.9%)	60 (69.0%)	77 (88.5%)	53 (60.9%)	25 (28.7%)	25 (28.7%)
<b>Total classes associated</b>	04	14	27	10	34	62	62
– Appropriate associations (True Positive)	04 (4.6%)	14 (16.1%)	26 (29.9%)	09 (10.4%)	32 (36.8%)	62 (71.3%)	60 (69.0%)
– Inappropriate associations (False Positive)	0	0	01 (1.1%)	01 (1.1%)	02 (2.3%)	0	02 (2.3%)

It is noteworthy that while omission rates were high for traditional models, commission errors (false positives) were negligible across all versions (<3%) – Table 2 and Figure 10. This suggests that current LLMs adopt a conservative approach to alignment: they prefer to omit an association rather than generate an incorrect one. The few errors observed involved generic mappings (e.g., associating 'Mineral Extraction' broadly with building=industrial), which, while technically imprecise, are not semantically incoherent.



**Figure 10.** Comparative performance of semantic recall. The visual analysis highlights the significant reduction in omission errors (grey bars) achieved by reasoning-oriented models compared to traditional architectures.

## 5.2. Syntactic Conformity and Structural Integrity

While semantic retrieval improved, the ability to generate valid, structurally sound OWL code remains a significant bottleneck (Table 3). DeepSeek and early Gemini version failed to preserve the input hierarchy, frequently inverting relationships by making the original EDGV classes subclasses of OSM tags. This fundamental logical error highlights the distinction between linguistic understanding (identifying that ‘Hospital’ relates to ‘Health’) and formal reasoning (understanding that a Class cannot logically be a subclass of an Attribute).

ChatGPT demonstrated the highest syntactic stability, creating valid ‘EquivalentTo’ axioms and preserving object properties. Conversely, Gemini 2.5 Pro, despite its superior semantic recall, altered original class names (e.g., merging keys into names like OSM\_key\_value), compromising the ontology’s integrity. These results align with those of [35], who observed that LLMs struggle to self-correct in code-generation tasks, often prioritizing content over syntax.

**Table 3.** Syntactic conformity and structural integrity of generated ontologies.

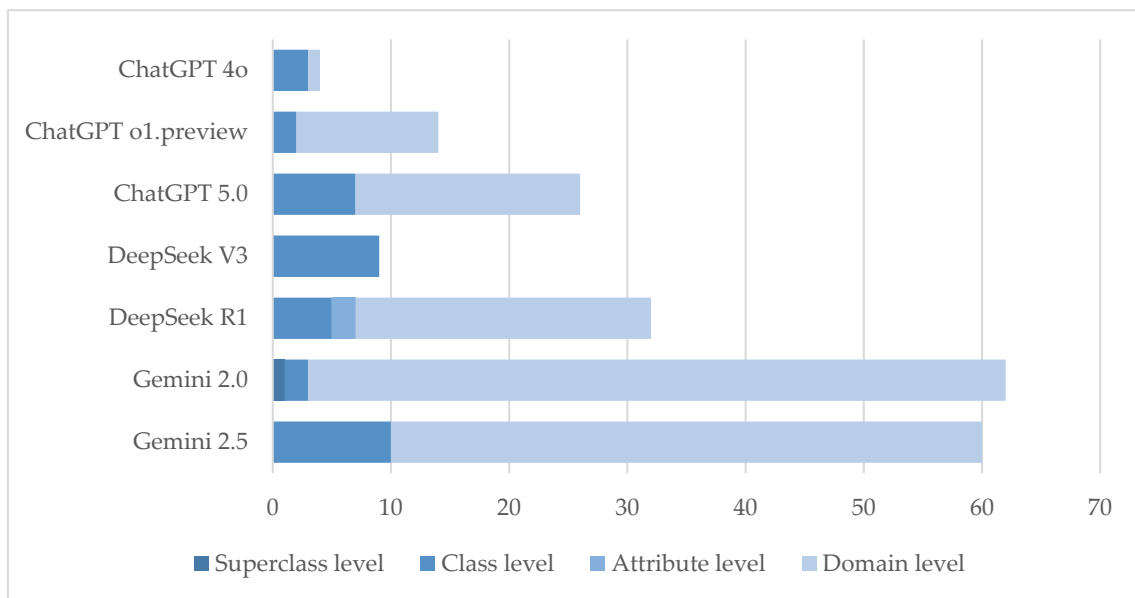
Model Family & Version	Generated Tags (n)	Structural Integrity (Hierarchy & Classes)	Naming Convention	Ontology Components Preservation (Props/Instances)	Mapping Logic (Association Method)
<b>OpenAI</b>					
ChatGPT 4o	4	<b>Failed.</b> Lost hierarchy; retained only associated classes (flat structure).	key=value	<b>Low.</b> Lost properties and instances.	EquivalentTo + ObjectProperty (hasOSMTag)
ChatGPT o1-preview	14	<b>Partial.</b> Retained associated classes but lost global hierarchy.	key_value	<b>Medium.</b> Retained properties; lost annotations/instances.	EquivalentTo + ObjectProperty (mapsTo...)
ChatGPT 5.0	31	<b>High.</b> Preserved full hierarchy and original classes.	OSM_key_value	<b>High.</b> Retained annotations and properties. Instances unlinked.	EquivalentTo + ObjectProperty (associated_with)
<b>DeepSeek</b>					
DeepSeek V3	11	<b>Failed (Inverted).</b> Created OSM superclasses containing EDGV subclasses.	key	<b>Low.</b> Lost all components.	SubclassOf + ObjectProperty
DeepSeek R1	34	<b>Failed.</b> Treated original classes as subclasses of OSM tags.	OSM_key_value	<b>Low.</b> Lost all components.	EquivalentTo (direct naming association)
<b>Google</b>					
Gemini 2.0 Flash	62	<b>Failed (Inverted).</b> Similar to V3; inverted hierarchy structure.	OSM_Tags_key_value	<b>Low.</b> Lost all components.	SubclassOf (direct)
Gemini 2.5 Pro	62	<b>High (Renamed).</b> Preserved hierarchy but renamed original classes.	key_value = Class	<b>High.</b> Preserved annotations, properties, and instances.	Mixed: EquivalentTo and SubclassOf (Inconsistent)

### 5.3. Reasoning Depth: Hierarchical Levels and Multiplicity

Cognitive theory posits that matching concepts at higher levels of abstraction (Classes) are more cognitively demanding than matching direct instances or domains [1,36]. The data in Table 4 and Figure 11 support this: traditional models predominantly formed associations at the simpler Domain level (direct term translation). In contrast, reasoning models (DeepSeek R1, Gemini 2.5) increasingly operated at the Class level, demonstrating an ability to infer broader categorical equivalence.

**Table 4.** EDGV taxonomy level with semantic associations.

Taxonomic Level	ChatGPT			DeepSeek		Gemini	
	4o	o1	5.0	V3	R1	2.0	2.5
<i>Unassociated classes (Omission)</i>	83	73	60	77	53	25	25
<i>Appropriate associations (Total)</i>	<b>4</b>	<b>14</b>	<b>26</b>	<b>9</b>	<b>32</b>	<b>62</b>	<b>60</b>
– Superclass level	0	0	0	0	0	1	0
– Class level	3	2	7	9	5	2	10
– Attribute level	0	0	0	0	2	0	0
– Domain level	1	12	19	0	25	59	50

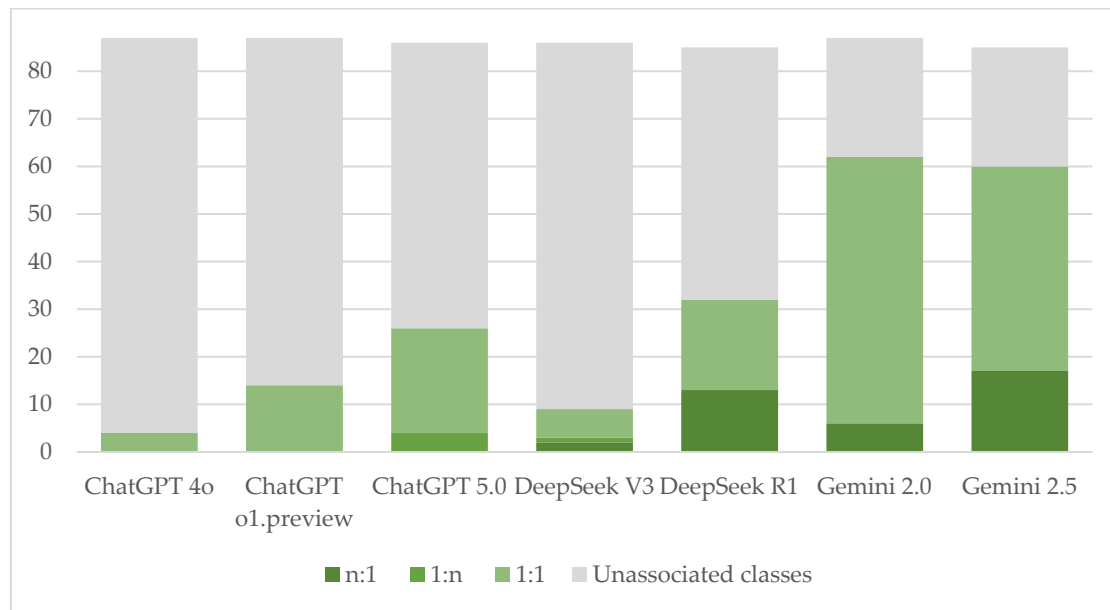


**Figure 11.** Analysis of reasoning depth by taxonomic level. While traditional models rely predominantly on lexical matching at the ‘Domain’ level, Large Reasoning Models (LRMs) demonstrate advanced inference capabilities, identifying a significantly higher number of associations at the abstract ‘Class’ level.

This reasoning capacity is further evidenced by the handling of multiplicity (Table 5 and Figure 12). Semantic alignment is rarely one-to-one; it often requires aggregating multiple source concepts into a single target tag (n:1).

**Table 5.** Semantic associations categorized by mapping multiplicity.

Mapping Multiplicity	ChatGPT			DeepSeek		Gemini	
	4o	o1	5.0	V3	R1	2.0	2.5
<i>Unassociated classes (Omission)</i>	83	73	60	77	53	25	25
<i>Appropriate associations (Total)</i>	<b>4</b>	<b>14</b>	<b>26</b>	<b>9</b>	<b>32</b>	<b>62</b>	<b>60</b>
– 1:1 mapping (One-to-One)	4	14	22	6	19	56	43
– 1:n mapping (One-to-Many)	0	0	4	1	0	0	0
– n:1 mapping (Many-to-One)	0	0	0	2	13	6	17



**Figure 12.** Complexity of semantic mappings. The chart details the cardinality of the associations identified by each model. While traditional models are limited to simple one-to-one (1:1) pairings, reasoning models can infer complex many-to-one (n:1) relationships, aggregating multiple specific domains into broader OSM tags.

While traditional models defaulted to simple 1:1 mappings, reasoning models successfully identified complex n:1 relationships. For instance, Gemini 2.5 correctly aggregated diverse EDGV religious domains ('Church', 'Mosque', 'Temple') into a single OSM tag, `amenity=place_of_worship`. This capability indicates a shift from mere lexical matching to genuine semantic inference, addressing a core challenge in aligning authoritative taxonomies with collaborative folksonomies [6].

#### 5.4. Theoretical Implications and Comparison with Related Work

The findings reveal a heterogeneous landscape where no single model achieved consistent success across syntactic validity, semantic coherence, and hierarchical interpretation. This partially contrasts with earlier studies that reported promising results for LLMs in purely textual geospatial tasks, such as identifying lexical similarities or interpreting map legends. Our results demonstrate that transitioning from unstructured text to formal OWL ontologies introduces a "structural gap" that current models struggle to bridge autonomously.

This difficulty aligns with broader AI research, indicating that while LLMs excel at natural language reasoning, they face significant constraints with logical consistency and rule-based systems [16–18]. The frequent syntactic errors observed in OWL generation, inverting hierarchies or corrupting OWL notation, mirror documented weaknesses in formal code generation. However, the discernibly better performance of Reasoning Models (LRMs) in preserving subclass structures supports the "neuro-symbolic" proposition. By using the ontology as a formal scaffold, these models achieved a higher level of coherence than previous unstructured approaches, suggesting that a hybrid workflow, when combining LLM semantic suggestions with LRM structural checks, is the most viable path for operational interoperability.

#### 5.5. Limitations and Directions for Future Research

This study presents constraints that outline clear avenues for future investigation. Firstly, the evaluation focused on a single EDGV category ('Buildings'); future work should expand to diverse domains and to other standards, such as CityGML or INSPIRE, to test generalizability. Secondly, the reliance on manual evaluation, while necessary for interpretive accuracy, introduces subjectivity. Subsequent studies should integrate automated reasoning engines (e.g., HermiT or Pellet) to validate logical consistency at scale. Finally, we employed a uniform prompt strategy to ensure comparability.

Future research must therefore explore advanced prompt engineering techniques, such as chain-of-thought or schema-specific constraints, to determine if syntactic limitations can be mitigated through procedural guidance rather than architectural changes alone.

As noted by [37], true understanding demands a grasp of the implicit rules governing how humans organize reality. To address this, future research must move beyond descriptive semantics to address the 'spatial reasoning gap'. A critical next step is to incorporate geometric primitives and topological relationships into the ontological reasoning process, achieving the understanding of the entire semantic structure.

## 6. Conclusions

This study reveals that Large Language Models, when grounded by formal ontologies, possess the capability to bridge the semantic gap between authoritative schemas (EDGV) and collaborative folksonomies (OSM), validating the 'neuro-symbolic' hypothesis. While traditional LLMs rely on superficial pattern matching, reasoning-oriented models (LRMs) guided by an ontological structure simulate complex semantic inference, effectively identifying equivalences across linguistic barriers. In this context, the ontology served as a crucial semantic and structural scaffold, guiding the models beyond purely linguistic similarity and enabling partial reasoning over class hierarchies.

Consequently, the results indicate that hybrid workflows combining ontology-encoded knowledge, LLM-based semantic proposals, LRM-based structural reasoning, and expert validation represent a practical approach to advancing semantic interoperability. While current models cannot yet replace formal methods, they can generate valuable first drafts and reveal inconsistencies, accelerating the early stages of alignment design in Spatial Data Infrastructures.

However, significant limitations persist. The models struggled with syntactic generation, highlighting a disconnect between semantic comprehension and logical structuring. Cognitively, this underscores the challenge of automating 'tacit knowledge'. While LRMs inferred complex relationships, they lack the genuine spatial perception required to fully understand cartographic generalization and scale. Enhancing LLMs to process not only feature names but also their shapes and spatial contexts will be essential to achieving fully automated, robust semantic interoperability in the next generation of Geospatial AI.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

**Data Availability Statement:** The data and materials that support the findings of this study are openly available. The repository includes: (1) the reference EDGV ontology (OWL); (2) the complete set of ontologies generated by the evaluated LLMs/LRMs; (3) all figures in high resolution; and (4) a detailed spreadsheet containing the full list of semantic associations and the data used in the graphs. These resources can be accessed at: [https://anonymous.4open.science/r/reasoning\\_ontology-C795/](https://anonymous.4open.science/r/reasoning_ontology-C795/).

**Conflicts of Interest:** The authors report there are no competing interests to declare.

## Abbreviations

The following abbreviations are used in this manuscript:

EDGV	Estruturação de Dados Geoespaciais Vetoriais (Brazilian Portuguese)
LLM	Large Language Model
LRM	Large Reasoning Model
OSM	OpenStreetMap
OWL	Ontology Web Language

## Notes

1. Access to ChatGPT: <https://openai.com/index/chatgpt/>

2. Access to DeepSeek: <https://www.deepseek.com/>
3. Access to Gemini: <https://gemini.google.com/app>

## References

1. Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192–233. <https://doi.org/10.1037/0096-3445.104.3.192>.
2. Rosch, E. (1973). Natural categories. *Cognitive Psychology*, vol.4.
3. Petchenik, B. B. (1977). Cognition In Cartography. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 14(1), 117–128. <https://doi.org/10.3138/97R4-84N4-4226-0P24>.
4. Fremlin, G., & Robinson, A. H. (1998). What Is It That Is Represented on a Topographical Map? *Cartographica: The International Journal for Geographic Information and Geovisualization*, 35(1-2), 13-19.
5. Janowicz, K.; Scheider, S.; Adams, B. (2013). A Geo-semantics Flyby. In: (Eds) Rudolph S.; Gottlob G.; Horrocks, I.; Van Harmelen; F. Reasoning Web: Semantic Technologies for Intelligent Data Access. Reasoning Web 2013. Lecture Notes in Computer Science, v. 8067. Springer, Berlin, Heidelberg, p. 230-250.
6. Machado, A. A., & Camboim, S. P. (2024). Semantic Alignment of Official and Collaborative Geospatial Data: A Case Study in Brazil. *Revista Brasileira de Cartografia*, 76. Scopus. <https://doi.org/10.14393/rbcv76n0a-72070>.
7. Souza, F. A., da Silva, E. D. B., & Camboim, S. P. (2025). Explorando o Uso de Large Language Model (ChatGPT) para Alinhamento Semântico entre Esquemas Conceituais de Dados Geoespaciais. *Rev. Bras. Cartogr*, 77, 1.
8. Souza, F. A., & Camboim, S. P. (2024). Advancing Geospatial Data Integration: The Role of Prompt Engineering in Semantic Association with chatGPT. *Free and Open-Source Software for Geospatial 2024 (FOSS4G 2024)*, Belém,PA,Brazil, 02-08-December, 2024 (Session Academic Track, Part Full Papers, pp 87-92). <https://zenodo.org/records/14250739>
9. Souza, F. A., & Camboim, S. P. (2023). Semantic Alignment of Geospatial Data Models using chatGPT: preliminary studies. In da Fonseca Feitosa F. & Vinhas L. (Eds.), *Proc. Brazilian Symp. GeoInformatics* (pp. 399–404). National Institute for Space Research, INPE; Scopus. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85181118913&partnerID=40&md5=45de9b24f4242bc1e4306f46b84a1ed0>
10. Anand, S., Morley, J., Jiang, W., Du, H., & Hart, G. (2010). When worlds collide: Combining Ordnance Survey and Open Street Map data. In: *AGI Geocommunity '10*, London, UK.
11. Ballatore, A., Bertolotto, M., & Wilson, D. C. (2013). Geographic knowledge extraction and semantic similarity in OpenStreetMap. *Knowledge and Information Systems*, 37(1), 61–81. <https://doi.org/10.1007/s10115-012-0571-0>
12. Du, H., Alechina, N., Jackson, M., & Hart, G. (2013). Matching Formal and Informal Geospatial Ontologies. In D. Vandenbroucke, B. Bucher, & J. Crompvoets (Eds.), *Geographic Information Science at the Heart of Europe* (pp. 155–171). Springer International Publishing. [https://doi.org/10.1007/978-3-319-00615-4\\_9](https://doi.org/10.1007/978-3-319-00615-4_9)
13. Yu, L., Qiu, P., Liu, X., Lu, F., & Wan, B. (2018). A holistic approach to aligning geospatial data with multidimensional similarity measuring. *International Journal of Digital Earth*, 11(8), 845–862. <https://doi.org/10.1080/17538947.2017.1359688>
14. Romanenko, E., Calvanese, D., & Guizzardi, G. (2024). Evaluating quality of ontology-driven conceptual models abstractions. *Data & Knowledge Engineering*, 153, 102342. <https://doi.org/10.1016/j.datak.2024.102342>
15. Kang, Y., Gao, S., & Roth, R. (2024). Artificial intelligence studies in cartography: A review and synthesis of methods, applications, and ethics. *Cartography and Geographic Information Science*. <https://doi.org/10.1080/15230406.2023.2295943>
16. Mirzadeh, I., Alizadeh, K., Shahrokhi, H., Tuzel, O., Bengio, S., & Farajtabar, M. (2024). GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models (arXiv:2410.05229). arXiv. <https://doi.org/10.48550/arXiv.2410.05229>

17. Tucker, S. (2024). A systematic review of geospatial location embedding approaches in large language models: A path to spatial AI systems. arXiv, Jan. 12, 2024. doi: 10.48550. arXiv preprint arXiv.2401.10279, 10.
18. Valmeekam, K., Stechly, K., & Kambhampati, S. (2024). LLMs Still Can't Plan; Can LRMs? A Preliminary Evaluation of OpenAI's o1 on PlanBench (arXiv:2409.13373). arXiv. <https://doi.org/10.48550/arXiv.2409.13373>
19. Concar. (2017). Comissão Nacional de Cartografia. Especificações Técnicas para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0). NCB-CC/E 0001B08. Versão 3.0.
20. Grinberger, A. Y., Minghini, M., Juhász, L., Yeboah, G., & Mooney, P. (2022). OSM Science - The Academic Study of the OpenStreetMap Project, Data, Contributors, Community, and Applications. *ISPRS International Journal of Geo-Information*, 11(4), 230. <https://doi.org/10.3390/ijgi11040230>
21. Kaur, J., Singh, J., Sehra, S. S., & Rai, H. S. (2017). Systematic Literature Review of Data Quality Within OpenStreetMap. 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS), 177–182. <https://doi.org/10.1109/ICNGCIS.2017.35>
22. OSM. (2025). OpenStreetMap: Map Features. [https://wiki.openstreetmap.org/wiki/Map\\_features](https://wiki.openstreetmap.org/wiki/Map_features)
23. Liang, B., Wang, Y., & Tong, C. (2025). AI Reasoning in Deep Learning Era: From Symbolic AI to Neural-Symbolic AI. *Mathematics*, 13(11), 1707. <https://doi.org/10.3390/math13111707>
24. Mira, J. M. (2008). Symbols versus connections: 50 years of artificial intelligence. *Neurocomputing*, 71(4-6), 671-680.
25. Santhanam, S.; Shaikh, S. (2019). A Survey of Natural Language Generation Techniques with a Focus on Dialogue Systems - Past, Present and Future Directions. Cornell University. <https://arxiv.org/abs/1906.00500>
26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
27. Jozefowicz, R.; Vinyals, O.; Schuster, M.; Shazeer, N.; Wu, Y. (2016). Exploring the Limits of Language Modeling. <https://arxiv.org/abs/1602.02410>
28. Prince, S. J. D. (2023). Understanding Deep Learning. <http://udlbook.com>.
29. Berners-Lee, T.; Hendler, J.; Lassila, O. (2001). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, a division of Nature America, Inc. Vol. 284, No. 5, pp. 34-43.
30. Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. <https://doi.org/10.1006/knac.1993.1008>
31. Zhang, Y., Wei, C., He, Z., & Yu, W. (2024). GeoGPT: An assistant for understanding and processing geospatial tasks. *International Journal of Applied Earth Observation and Geoinformation*, 131. Scopus. <https://doi.org/10.1016/j.jag.2024.103976>
32. Mooney, P., Cui, W., Guan, B., & Juhász, L. (2023). Towards Understanding the Geospatial Skills of ChatGPT Taking a Geographic Information Systems (GIS) Exam. In S. Newsam, L. Yang, G. Mai, B. Martins, D. Lunga, & S. Gao (Eds.), *Maynooth University* (WOS:001152316700015; pp. 85–94). <https://doi.org/10.1145/3615886.3627745>
33. Borges, K. A. V., Davis Jr., C. A. & Laender, A. H. F. (2001). OMT-G: An Object-Oriented Data Model for Geographic Applications. *GeoInformatica* 5, 221–260. <https://doi.org/10.1023/A:1011482030093>
34. International Organization for Standardization. (2013). Geographic information – Data quality (ISO Standard No. 19157:2013[2014]). <https://www.iso.org/standard/32575.html>
35. Zhang, Q., Zhang, T., Zhai, J., Fang, C., Yu, B., Sun, W., & Chen, Z. (2024). A Critical Review of Large Language Model on Software Engineering: An Example from ChatGPT and Automated Program Repair. arXiv. <http://arxiv.org/abs/2310.08879>

36. Bravo, J. V. M. (2014). A Confiabilidade Semântica das Informações Geográficas Voluntárias como Função da Organização Mental do Conhecimento Espacial. Dissertação de Mestrado. 139 p. Universidade Federal do Paraná, Programa de Pós-Graduação em Ciências Geodésicas, Curitiba (PR).
37. McCarthy, J. (2007). What is Artificial Intelligence? Computer Science Department / Stanford University. <http://www-formal.stanford.edu/jmc/>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.