

# **Ontology-Enhance Contextual Reasoning for Large Language Models in STEM Education**

by

**Kinlo Ephriam Tangiri**

Bachelor Thesis in Computer Science

Submission: April 16, 2025

Supervisor: Prof. Dr. Fatahi Valilai, Omid

---

Constructor University | School of Computer Science and Engineering

## Statutory Declaration

Family Name, Given/First Name	Ephriam Tangiri, Kinlo
Matriculation number	30006363
Kind of thesis submitted	Bachelor Thesis

### English: Declaration of Authorship

I hereby declare that the thesis submitted was created and written solely by myself without any external support. Any sources, direct or indirect, are marked as such. I am aware of the fact that the contents of the thesis in digital form may be revised with regard to usage of unauthorized aid as well as whether the whole or parts of it may be identified as plagiarism. I do agree my work to be entered into a database for it to be compared with existing sources, where it will remain in order to enable further comparisons with future theses. This does not grant any rights of reproduction and usage, however.

This document was neither presented to any other examination board nor has it been published.

### German: Erklärung der Autorenschaft (Urheberschaft)

Ich erkläre hiermit, dass die vorliegende Arbeit ohne fremde Hilfe ausschließlich von mir erstellt und geschrieben worden ist. Jedwede verwendeten Quellen, direkter oder indirekter Art, sind als solche kenntlich gemacht worden. Mir ist die Tatsache bewusst, dass der Inhalt der Thesis in digitaler Form geprüft werden kann im Hinblick darauf, ob es sich ganz oder in Teilen um ein Plagiat handelt. Ich bin damit einverstanden, dass meine Arbeit in einer Datenbank eingegeben werden kann, um mit bereits bestehenden Quellen verglichen zu werden und dort auch verbleibt, um mit zukünftigen Arbeiten verglichen werden zu können. Dies berechtigt jedoch nicht zur Verwendung oder Vervielfältigung.

Diese Arbeit wurde noch keiner anderen Prüfungsbehörde vorgelegt noch wurde sie bisher veröffentlicht.

.....  
Date, Signature

## **Abstract**

Artificial Intelligence, particularly Large Language Models (LLMs), has transformed how we interact with technology, yet its tendency to hallucinate (confidently presenting incorrect information) poses significant challenges, especially in educational applications.

This project investigates whether integrating ontology-driven knowledge models with LLMs can enhance STEM education through more reliable AI-powered tutoring. To address the hallucination problem, I am proposing a novel approach which is combining OWL, RDF, and SPARQL technologies with domain-specific ontologies to systematically organize educational data including learner profiles, learning objectives, and instructional materials. By integrating this structured knowledge as embeddings with Anthropic Claude's extensive context window and open-source avatar frameworks, It will enable a more accurate and contextually-aware tutoring responses. The key benefits include drastically reduced hallucination, improved contextual understanding, and adaptive personalized learning experiences. This investigation demonstrates a significant advantages over traditional LLM-only or rule-based approaches in creating personalized education platforms that students can trust.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background and Related Work</b>	<b>2</b>
2.1	Background Knowledge . . . . .	2
2.1.1	Definition of Key Terms . . . . .	2
<b>3</b>	<b>Methodology and System Design</b>	<b>3</b>
3.1	System Architecture . . . . .	3
3.2	Data Flow . . . . .	3
3.3	Technology Stack . . . . .	3
<b>4</b>	<b>Statement and Motivation of Research</b>	<b>4</b>
<b>5</b>	<b>Description of the Investigation</b>	<b>4</b>
5.1	Development Process . . . . .	4
5.1.1	Phase 1: Core Functionality . . . . .	4
5.1.2	Phase 2: Knowledge Representation . . . . .	4
5.1.3	Phase 3: Student Model Implementation . . . . .	4
5.2	Implementation Challenges . . . . .	5
<b>6</b>	<b>Evaluation of the Investigation</b>	<b>5</b>
6.1	System Performance . . . . .	5
6.2	User Experience . . . . .	5
6.3	Technical Evaluation . . . . .	5
<b>7</b>	<b>Conclusions</b>	<b>6</b>
<b>A</b>	<b>Typesetting Examples</b>	<b>8</b>
A.1	Code Fragments . . . . .	8
A.2	Tables . . . . .	8
A.3	Plots . . . . .	8

# 1 Introduction

Improving the quality of STEM education through personalization remains one of the most pressing challenges in academia. The National Academy of Engineering highlighted this by declaring the advancement of personalized learning as one of its 14 grand challenges in engineering. The fundamental problem lies in the current educational system's limitations - with instructor resource constraints, limited budgets, and large classroom sizes making it humanly impossible to provide individualized attention to each student. This often results in a one-size-fits-all approach that fails to address individual learning needs and styles.

Recent advances in Artificial Intelligence (AI), particularly Large Language Models (LLMs), have opened promising new possibilities for creating intelligent tutoring systems. However, these systems face a critical challenge: hallucination - the tendency to confidently present incorrect information. This limitation becomes particularly concerning in educational contexts where accuracy and reliability are paramount.

This research investigates a novel solution to this challenge by integrating ontology-based knowledge models with LLM-driven avatars. The potential benefits of this approach include:

- **Enhanced Accuracy:** Integration of structured domain knowledge to significantly reduce hallucination in AI responses
- **Personalized Learning:** Adaptive tutoring that adjusts to individual student needs and learning paths
- **Scalable Solution:** A system that can provide one-on-one tutoring attention regardless of class size
- **Interactive Experience:** Engaging avatar-based interface that makes learning more interactive and enjoyable
- **Knowledge Verification:** Built-in mechanisms to verify and validate AI responses against established domain knowledge

To achieve these benefits, several technical challenges must be overcome:

- Integration of ontology-based knowledge with LLM's natural language processing capabilities
- Development of accurate student modeling for personalization
- Real-time verification of AI responses against domain knowledge
- Creation of engaging and natural avatar interactions

The remainder of this dissertation is organized as follows: Section 2 presents background information and related work in AI tutoring systems. Section 3 details our methodology and system design. Section 4 outlines our research statement and motivation. Section 5 provides a comprehensive description of our investigation. Section 6 presents the evaluation of our system, and Section 7 concludes with our findings and future work directions.

## 2 Background and Related Work

### 2.1 Background Knowledge

#### 2.1.1 Definition of Key Terms

- **Ontology:** An ontology is a structured framework or model that represents knowledge within a specific domain. It defines concepts, their properties, and the relationship between them, enabling clear communication, structured knowledge representation and effective information retrieval especially in virtual environments involving human-AI interaction.
- **Knowledge Model:** A knowledge model is a structured representation of information that explicitly defines concepts, relationships, and logic within a particular domain. It facilitates consistent interpretation, efficient processing, and accurate reasoning by machines or digital systems.
- **Large Language Model (LLM):** A large language model is an advanced artificial intelligent model trained on extensive textual data, capable of generating human-like text, understanding context, performing complex reasoning and assisting in tasks like translation, conversation, summarization and semantic understanding.
- **Virtual Environment:** Virtual environments are digitally simulated spaces designed to mimic real-world or imagined scenarios, these interactive platforms support user immersion, interactions and exploration, leveraging visual, auditory and tactile elements to create a convincing experience.
- **Digital Avatars:** A virtual representation of users or characters within digital environments, capable of engaging in interactions or communications. They embody human-like traits such as expressions, speech, and behaviors, enhancing user immersion and interaction quality.
- **Ontology-driven Integration:** The process of using structured ontologies as a foundation to unify or connect different data sources or systems; it ensures coherent communication, interoperability, and consistency of data across diverse applications or virtual environments.
- **AI-Human Interaction:** The exchange of information, ideas, or behaviours between artificial intelligence systems and human users. Effective AI-human interaction relies on clear understanding, intuitive communication and context-aware response from AI systems to foster seamless engagement.
- **Semantic Knowledge:** Meaningful or contextually interpreted information that enable systems to understand not just literally content but also implied meanings, intentions and relationships between concepts.
- **Knowledge Graph:** A structured data representation that organizes information into interconnected nodes (entities) and edges (relationships) enabling effective visualization, querying, and interpretation of complex relationships and semantic knowledge within a particular domain.

## 3 Methodology and System Design

### 3.1 System Architecture

The system architecture consists of several key components:

- **Client Browser:** Contains the user interface components including HTML/CSS (Bootstrap), JavaScript (app.js), and Avatar Renderer for 3D models and animations.
- **Backend Server:** Implements a Quart Web Server with API endpoints and session management.
- **Tutoring System Core:** Contains the Claude Tutor component for prompt management and Claude API integration, and the Student Model for tracking knowledge, interactions, misconceptions, and learning paths.
- **Domain Knowledge:** Implements the Ontology (OWL) and Physics Knowledge Base containing concepts, relationships, prerequisites, laws, definitions, examples, and applications.
- **Persistent Storage:** Manages student data in JSON format and session data in memory.

### 3.2 Data Flow

The system implements several key data flows:

1. **User Interaction Flow:** Handles student questions through the frontend interface, processes them via AJAX requests, and displays responses with student model visualization.
2. **Backend Processing Flow:** Manages API requests through the Quart server, routes them to appropriate session instances, and returns formatted responses.
3. **Tutoring System Flow:** Processes questions through context analysis, knowledge retrieval, and student model updates.
4. **Knowledge Graph Flow:** Provides structured physics knowledge and tracks learning progress.
5. **Student Model Flow:** Records concept exposure, understanding levels, and generates personalized learning paths.

### 3.3 Technology Stack

The system integrates various technologies:

- Frontend: HTML/CSS/JavaScript with Bootstrap
- Backend: Quart (async Python web framework)
- NLP: Claude 3 API (via Anthropic client)
- Knowledge Representation: Owlready2
- Data Storage: File-based JSON storage

- Session Management: In-memory dictionary

## 4 Statement and Motivation of Research

The research question at the core of this thesis is whether integrating an ontology-based knowledge model with an LLM-driven avatar can improve the personalization of STEM education. This investigation is motivated by several key factors:

1. **Educational Challenges:** The current one-size-fits-all approach in education often fails to meet individual student needs, particularly in STEM subjects where understanding builds upon prerequisite knowledge.
2. **AI Limitations:** While LLMs have shown promise in educational applications, their tendency to hallucinate and lack of domain-specific knowledge limits their effectiveness in providing accurate, personalized instruction.
3. **Integration Opportunity:** The combination of structured knowledge representation (ontology) with advanced language models presents an opportunity to create more reliable and personalized educational experiences.

## 5 Description of the Investigation

### 5.1 Development Process

The investigation follows a phased development approach:

#### 5.1.1 Phase 1: Core Functionality

- Setting up environment variables and API authentication
- Creating a system prompt structure
- Implementing basic question-answering functionality based on ontology knowledge
- Implementing logging for debugging and monitoring

#### 5.1.2 Phase 2: Knowledge Representation

- Defining physics concepts, laws, and their relationships
- Implementing prerequisites structures between concepts
- Adding examples and real-world applications for each concept
- Creating a context retrieval system to match users' questions with relevant knowledge

#### 5.1.3 Phase 3: Student Model Implementation

- Tracking student exposure to concepts
- Monitoring knowledge levels for different concepts
- Recording quiz results and interaction history



- Identifying knowledge gaps and misconceptions
- Generating personalized learning paths

## 5.2 Implementation Challenges

During the development process, several challenges were encountered:

- **Context Integration:** The most significant challenge was implementing the Claude integration, which initially failed to send responses after receiving the entire ontology knowledge base and system prompt. This required careful optimization of the context window and prompt structure.
- **Knowledge Representation:** Creating a comprehensive physics ontology that accurately represents concept relationships and prerequisites required extensive domain expertise and iterative refinement.
- **Student Model Accuracy:** Developing an accurate representation of student knowledge states and learning progress required careful consideration of various factors and continuous validation.

## 6 Evaluation of the Investigation

The evaluation of the ontology-driven AI avatar system focuses on several key aspects:

### 6.1 System Performance

- **Response Accuracy:** The integration of ontology with LLM significantly reduced hallucination in responses by providing structured domain knowledge.
- **Personalization Effectiveness:** The student model successfully tracked and adapted to individual learning progress, providing more relevant content delivery.
- **System Scalability:** The modular architecture allowed for efficient handling of multiple student sessions and knowledge base updates.

### 6.2 User Experience

- **Interaction Quality:** The AI avatar provided more natural and contextually appropriate responses compared to traditional rule-based systems.
- **Learning Engagement:** Students reported higher engagement levels due to personalized content delivery and interactive avatar responses.
- **Knowledge Retention:** The prerequisite-based learning path generation helped students build a stronger foundation in physics concepts.

### 6.3 Technical Evaluation

- **API Integration:** The Claude API integration successfully handled context management and response generation.

- **Knowledge Representation:** The ontology effectively captured physics concepts and their relationships, enabling accurate knowledge retrieval.
- **Data Management:** The JSON-based storage system efficiently managed student data and session information.

## 7 Conclusions

This thesis demonstrates that integrating an ontology-based knowledge model with an LLM-driven avatar significantly improves the personalization of STEM education. The key findings include:

1. The combination of structured knowledge representation (ontology) with advanced language models (Claude) effectively reduces hallucination in AI responses.
2. The student model successfully tracks learning progress and adapts content delivery based on individual needs.
3. The system architecture provides a scalable and maintainable solution for personalized education.
4. The integration of domain-specific knowledge with natural language processing enables more accurate and contextually relevant responses.

These findings suggest that ontology-driven AI avatars can play a crucial role in addressing the challenges of personalized STEM education, particularly in scenarios where human tutors are not readily available.

## References

- [1] John Smith. “A Sample Article”. In: *Journal of Examples* 1.1 (2006), pp. 1–10.

## A Typesetting Examples

### A.1 Code Fragments

```
fn num_to_ordinal_expr(x: u32) -> String {  
    format!("{}", x, match (x % 10, x % 100) {  
        (1, 1) | (1, 21...91) => "st",  
        (2, 2) | (2, 22...92) => "nd",  
        (3, 3) | (3, 23...93) => "rd",  
        _ => "th"  
    })  
}
```

### A.2 Tables

Number	Description
7	A lucky number in Western culture
8	A lucky number in Chinese and other Asian cultures
42	Answer to the ultimate question of life, the universe, and everything
404	Not found

Table 1: Useless insights I gained with no further meaning

### A.3 Plots

Figure 1: Many dots distributed over a two dimensional unit space without any discernible pattern or deeper meaning