Utrecht University
Department of Information and Computing Science

Applied Data Science master thesis

Ontology-Based Generation of Medical Diagnostic Tasks and Feedback for Medical Students

First examiner:
Dr. Sergey Sosnovsky

Second examiner:
Dr. Hanna Hauptmann

Candidate:
Andrei Marinin

In cooperation with:
MLX Amsterdam company

July 12, 2024

Abstract

This thesis presents the design and evaluation of a prototype application aimed at enhancing medical education through the use of ontology-driven feedback. The application generates medical case studies that simulate real-life scenarios, focusing on diagnosing ear diseases. The pilot study, conducted with a small group of medical students, demonstrated the potential value of ontology-driven feedback in guiding students through the diagnostic process. Detailed log analysis identified common mistakes and knowledge gaps among students, highlighting areas where the prototype can be tailored to address these issues in future iterations. The learning process observed, with participants refining their choices and improving their accuracy, underscores the application's potential to enhance medical training. User experience feedback indicated that while the application was generally useful and informative, improvements in design and usability are needed to increase engagement and intuitiveness. Although the study did not show a significant improvement between pre- and post-test scores, the overall feedback from participants and their progression through exercises suggest that this approach effectively supports the learning process. The study faced limitations, including the specific focus on otology, time constraints in developing a user-friendly interface, and challenges in recruiting a sufficient number of medical students for evaluation. These factors highlight the need for future studies with larger participant pools to validate the findings and refine the application. Additionally, expanding the application's capabilities to cover other ENT diseases will broaden its utility. Writing scientific publications and presenting at international conferences are planned.

Keywords: ontology, medical education, interactive learning tools, ontology-driven feedback, learning exercise generation

Contents

1. Introduction

1.1 MLX Amsterdam Company

This thesis was written in cooperation with MLX Amsterdam Company (MLX). Founded in 2012, it focuses on creating interactive educational materials for medical students.

1.2 Motivation

The field of medical education stands at a crossroads where traditional learning methods often fail to meet contemporary demands. The motivation behind this thesis is rooted in the growing necessity for interactive and immersive educational tools in medical training [1]. As medical science evolves, so does the complexity of clinical scenarios that future healthcare professionals must be prepared to handle. Traditional educational approaches, such as lectures, textbook study, and seminars, while foundational, often fall short in replicating the dynamic and unpredictable nature of real-world medical situations [2]. This gap in medical education underscores the importance of developing training tools that not only convey theoretical knowledge but also simulate practical, hands-on experience. The development of a user-friendly interface prototype that can generate real-world medical exercises with detailed feedback presents a promising path to address these educational challenges and identify the knowledge gaps of medical students for the future enhancement of the learning process.

1.3 Problem Description

The core problem this thesis addresses is the lack of interactive and engaging training tools that accurately mimic real-world clinical scenarios. While digital medical education offers a plethora of theoretical exercises and tests, creating practical exercises for medical students remains a significant challenge [3].

Traditional methods like lectures and textbooks study in medical education primarily focus on theoretical knowledge, often overlooking the practical nuances and critical decision-making skills required in real-life medical practice. Current solutions in the field lack the ability to generate engaging medical exercises that mimic real-world clinical scenarios and provide effective feedback, addressing practical nuances and decision-making skills.

1.4 Scientific Novelty and Practical Value

The scientific novelty of this thesis lies in its approach of using ontology-driven content generation and providing ontology-driven feedback in medical practice training tools. While there have been individual efforts to utilize ontologies in educational applications, this research specifically focuses on hands-on medical exercises based on ontology-driven content. The practical value is evident in the development of an application that enhances the learning experience through interactive features and ensures the accuracy and relevance of medical content through structured ontology.

1.5 Thesis Goals

This thesis aims to design an application and an ontology that facilitate the generation of medical exercises and provide detailed feedback to medical students, enhancing their skills in diagnosing ear diseases and highlighting the students' knowledge gaps. The goal includes conducting a pilot study to evaluate the effectiveness of the system and developing a prototype that demonstrates the potential of this approach. Additionally, this thesis seeks to propose an interface version that increases medical student engagement and identifies areas for future interface development.

To address this, the thesis will explore the development, implementation, and evaluation of this application.

2. Literature Review

2.1 Automated Tests and Exercises Generation for Educational Purposes

Automated test, question, and exercise generation systems are designed to improve the learning process, enabling learners to evaluate their knowledge, quickly acquire new information, and effectively assess their understanding.

The primary objectives of automated question generation systems are linked to enhancing educational processes, with a distinct focus on assessing learners' knowledge, facilitating knowledge acquisition, and validating information. According to Kurdi et al. [4], a significant number of automated question generation efforts is directed towards creating assessment tools that evaluate learners' understanding and grasp of various subjects. This aligns with the broader educational goal of not only testing but also reinforcing learning through practice and feedback mechanisms. Also, these systems play a significant role in knowledge acquisition, where the generation of questions and exercises encourages deeper engagement with learning materials, thereby promoting better retention and understanding. Lastly, validation, although not as widely cited as assessment and knowledge acquisition, emerges as an aspect, particularly in confirming the accuracy and relevance of educational content.

Furthermore, these systems aim to address the challenges associated with manual test and exercise creation, such as the time-intensive and costly processes developers often face. Sahar A. El-Rahman and A. Zolait [5] demonstrate in their study that the advantages of automated test generation systems over traditional manual test creation methods are significant, offering considerable time savings and a reduction in potential biases or errors associated with manual question selection. By automating the generation of test papers, educators can devote more time to teaching and less to the logistical challenges of test preparation, ensuring that assessments are both rigorous and aligned with educational standards.

In a similar vein, Westphal [6] presents a framework for automatically generating a range of different exercise tasks in Haskell-I/O programming. The framework aims to alleviate the burdens of manual task creation, saving educators valuable time and reducing issues such as plagiarism in e-learning environments.

Another important aspect of automated exercise generation, as reviewed by Sahar A. El-Rahman and A. Zolait [5], is the necessity for exercises, questions, or tests to be shuffled randomly to avoid repetition and promote a broad assessment range. This randomization is critical for maintaining the integrity and fairness of the examination process, ensuring that each test is unique and reflective of the entire curriculum. The authors of this paper further highlight the importance of this feature in their Automated Test Paper Generation system, utilizing a shuffling algorithm that ensures the randomization of questions, thereby preventing repetition and promoting a comprehensive assessment approach.

2.1.1 Methods of Generating Questions

Every automated exercise generation system leverages diverse approaches in generating questions, tailoring techniques to the specificity and complexity of the content being tested. In their work, Yao, X., Bouma, G., Zhang, Y. [7] explore several foundational methods for question generation, including semantic, syntax, template, rule, and scheme-based approaches. The semantic method focuses on understanding and manipulating the underlying meanings of sentences, utilizing Minimal Recursion Semantics to generate questions that are semantically grounded. The syntax-based method manipulates the syntactic structure of sentences to create new questions, maintaining grammatical correctness. Template-based approaches rely on predefined patterns to produce questions, often requiring significant manual input but ensuring high relevance and accuracy. Rule-based methods apply specific linguistic or logical rules to transform statements into questions, offering a balance between automation and control. Lastly, scheme-based approaches use a combination of these methods, often integrating semantic and syntactic analyses to generate more complex and varied questions. This multi-faceted approach allows for the generation of questions that are not only grammatically correct but also meaningful and contextually appropriate, enhancing the learning experience through targeted assessment.

Building upon the insights provided by Yao and colleagues, Das, Majumder, Phadikar, and Sekh [8] further expand the discourse on automated tests generation techniques. They provide a comprehensive examination of automated question generation techniques, outlining them into rule-based, template-based, and artificial intelligence (AI)-based categories. Each method is explored for its unique advantages and inherent constraints within the educational technology landscape. Rule-based approaches, grounded in predefined rules and structures, are praised for their reliability in specific contexts but criticized for their lack of flexibility. Template-based methods offer a balance between structure and adaptability, utilizing templates to generate questions but may suffer from predictability and limited variability. AI-based techniques, particularly those leveraging recent advancements in machine learning and natural language processing, are recognized for their potential to generate more diverse and contextually rich questions.

2.2 Ontology for Knowledge Representation and for Learning Support

2.2.1 Using Ontologies for Knowledge Representation in the Medical Domain

Ontologies are used widely as a learning support in different domains, enhancing the precision and effectiveness of knowledge acquisition, management, and dissemination. Ontologies significantly enrich the representation and organization of knowledge, offering a structured framework that facilitates the understanding and processing of complex information across various domains, specifically in the medical domain [9].

The significance of ontologies extends into the realms of biological and biomedical research, where they are instrumental in providing standardized identifiers, vocabularies, metadata, and machine-readable axioms. Such tools are essential for the integrated analysis and interpretation of data, as highlighted by Hoehndorf, Schofield, and Gkoutos [10].

In medicine, ontologies enable the precise representation of clinical knowledge. For instance, Harispe et al. [11] propose a framework for ontology-based semantic similarity measures, emphasizing their application in various domains, including medicine, to enhance the accuracy and relevance of data analysis.

Interoperability emerges as a significant advantage of ontologies, allowing diverse systems to communicate and exchange information seamlessly. In the realm of biomedical research, Bodenreider and Stevens [12] reflect on the evolution of ontology in biomedicine, underscoring the move towards greater standardization and interoperability as a means to facilitate research and healthcare delivery.

Ontologies must be scalable to accommodate the ever-growing volume of medical data. Recent advancements, such as the work by Bo Liu et al. [13] on an incremental and distributed inference method for large-scale ontologies using MapReduce, address the scalability challenges by facilitating efficient and scalable reasoning in large ontology bases.

Disease-specific ontologies, like the Alzheimer's Disease Ontology discussed by Malhotra et al. [14], exemplify the application of ontologies in capturing domain-specific knowledge, enabling targeted research and personalized medicine.

Smith et al [15] provide a survey of biomedical imaging ontologies, outlining the challenges faced in fields like histopathological imaging and suggesting strategies for these challenges. This work highlights the importance of interoperability in integrating imaging data with other biomedical information.

The essence of ontologies lies in their ability to conceptualize domains through the definition of entities and the relationships between them. This is exemplified in the work of Zhang and Bodenreider [16], who elucidate the alignment of comprehensive anatomical ontologies, highlighting the semantic techniques critical for navigating the complexity inherent in such endeavors. This process is foundational for ensuring a coherent representation of domain-specific knowledge.

Ontologies contribute to the standardization of terminologies within domains, promoting consistency in the use of terms and definitions. Bard and Rhee [17] highlight the expansive applications of ontologies in biology, from gene annotation to the elucidation of disease mechanisms and the amalgamation of varied biological data. Ontologies are lauded for their capacity to connect disparate research areas, fostering novel insights and linkages that are essential for scientific advancement. This utility in synthesizing complex biological information underscores the pivotal role of ontologies in promoting interdisciplinary discoveries. Despite these benefits, Bard and Rhee also address the challenges confronting the development and application of ontologies, such as the intricacies of biological data, the ongoing need for updates and maintenance, and ensuring interoperability among different ontology systems.

2.2.2 Using Ontologies for Learning Support

Ontologies, as mentioned earlier, offer a structured content framework, enabling e-learning systems to deliver personalized experiences and manage content efficiently, as discussed by Sampson, Lytras, Wagner, and Diaz [18]. This editorial underscores the role of these technologies in facilitating content reuse and standardization, alongside generating adaptive learning paths tailored to individual learner needs.

One of the most vast domains of applying ontology for learning support is language studies. For example, Jingyun Wang et al. [19] illustrate the development and application of a course-centered ontology to enhance Japanese language learning. Their research focuses on constructing a comprehensive ontology-based system, specifically designed to map and visualize the complex interrelationships between various grammar points within a Japanese language course. This innovative approach not only facilitates a deeper understanding of language concepts by highlighting their interconnections but also demonstrates a significant improvement in learning outcomes. Through the use of this system, learners are encouraged to actively compare and contrast related language points, fostering a more integrated and holistic grasp of the language.

In the paper by Chen et al. [20], ontology serves as a technology for achieving the learning goals of users within an e-learning system designed for mobile phone salespersons. The ontology-based approach organizes domain-specific knowledge into a structured knowledge base. This structured organization enables the identification of personalized learning needs through an ontology-based quizzing module, directly addressing the individual knowledge gaps of the salespersons. As a result, the system can provide tailored learning content that specifically targets areas where each user lacks understanding, thereby effectively meeting their learning goals.

2.3 Role of Feedback in Automated Tests and Exercises Generation Systems

Feedback is conceptualized as information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding. Hattie and Timperley [21] identify several forms of feedback, such as corrective feedback from teachers or parents focusing on correcting mistakes, alternative strategy feedback from peers, clarification feedback from resources, encouragement feedback from parents or peers, and self-evaluation feedback where learners assess their own performance. Feedback aims to fill the gap between current understanding and desired goals, utilizing various processes like increasing effort, motivation, engagement, restructuring understandings, and confirming correctness. According to the study's findings, feedback is most effective when it addresses the task level (specific task information), process level (strategies and processes), and self-regulation level (self-monitoring and regulation). The authors also show that the most effective feedback provides cues or reinforcement, involves computer-assisted instructional feedback, and relates to specific goals.

Shute et al. [22] provide a comprehensive review highlighting several key aspects of formative feedback, emphasizing its delivery, types, timing, and the various variables that influence its effectiveness. The authors stress that the manner of feedback delivery plays a critical role in its ability to enhance learning, aligning with other studies that emphasize the importance of feedback quality in educational interventions [21].

Shute et al. categorize feedback into various types, such as verification of response accuracy, explanations of correct answers, and hints. Each type has a different impact on learning. For instance, verification provides immediate affirmation of correctness, while explanatory feedback aids deeper understanding. The timing of feedback is also pivotal; immediate feedback can prevent the reinforcement of errors, whereas delayed feedback can encourage deeper cognitive processing. The efficacy of immediate versus delayed feedback varies depending on the task complexity and the learner's proficiency [23].

The Shute et al. study identifies learner characteristics (e.g., prior knowledge and motivation), task nature, and specific feedback features as critical factors. These variables can lead to varying and sometimes inconsistent outcomes in feedback research. The evaluation of this study shows that feedback should focus on the task rather than the learner, be specific and clear, and be presented in manageable units. It should align with the learner's current understanding and ability level to be most beneficial. Furthermore, feedback should aim to reduce uncertainty between performance and goals and promote a learning goal orientation.

Mark Gierl and H. Lai [24] explore automatic item generation for computerized formative testing in medical education. Their study focuses on generating both test items and rationales, addressing the need for large numbers of diverse, high-quality test items in computerized testing. In addition, Mark Gierl and H. Lai in their study underscores the crucial role of feedback. The authors emphasize that feedback, particularly when integrated with automated tests and questions generation, becomes a tool for enhancing learning. Feedback in the context of automated exercise item generation is designed to provide learners with immediate, specific insights into both their correct and incorrect responses. Through detailed rationales that accompany each question, learners receive guidance on why an answer is correct or incorrect, which in turn fosters a deeper understanding of the material and encourages self-directed learning and critical thinking.

Mark Gierl and H. Lai in their research further demonstrates that effective feedback within computerized formative testing can significantly impact medical students' learning strategies. By aligning feedback with specific learning objectives, the automated generation system ensures that students not only assess their knowledge but also gain insights into their learning process, thereby enhancing the overall educational experience.

Another study that focuses both on the exercise generation and the role of feedback for learners is by Gütl et al. [25]. The development of the Enhanced Automatic Question Creator by Gütl and colleagues through its integration of natural language processing and AI, facilitates the automatic creation of test items from textual content. This system not only exemplifies the enhancement of learning experiences through diverse and personalized content but also highlights the capability for immediate feedback. The evaluation shows that immediate feedback mechanism is crucial for adapting teaching strategies and improving learners' knowledge acquisition and skills development.

There is an ontology-based feedback system, called OntoPeFeGe (Ontology-Based Personalized Feedback Generator), developed by Demaidi, M. N., et al. [26] to evaluate the role of feedback in educational settings. Ontology-based feedback in their work provides personalized, immediate, and formative responses to students. This approach adapts feedback to the learner's background knowledge and understanding level, ensuring relevance and efficiency. OntoPeFeGe exemplifies auto-generated feedback systems that dynamically create feedback using ontologies. This system integrates various feedback generation strategies and templates to provide comprehensive feedback based on Bloom's taxonomy levels. It employs a rule-based algorithm to tailor feedback according to student characteristics and assessment question difficulty, determining the appropriate type of feedback. Evaluations of OntoPeFeGe show that personalized feedback significantly improves the performance of students with lower background knowledge.

2.4 Learning Systems and Applications in Medical Education

The landscape of technological support for learning within medical education is vast and diverse. A multitude of solutions exists, ranging from augmented reality and virtual reality (Antoniou et al. [27]) to simulations and the generation of training exercises. In this research, the primary focus will be on simulations—particularly patient simulations—along with the generation of automated exercises and ontology-based learning support systems.

2.4.1 Ontology-Based Learning Support Systems in Medical Education

Leo et al. [28] present an ontology-based approach for generating complex multiple choice questions in medical education. These multiple choice questions are designed to simulate real-life medical scenarios. The authors exploit medical ontologies to generate case-based questions, demonstrating the practicality and effectiveness of their method in a user study with medical experts. Their system, EMMeT Multiple Choice Question Generator, uses a template-based approach, leveraging the Elsevier Merged Medical Taxonomy (EMMeT) ontology for content.

Radović, Petrovic, and Tosic [29] proposed an approach to medical education through SCTonto, an automated Script Concordance Test (SCT) generation platform. This ontology-driven learning assessment tool utilizes the SCT to evaluate clinical reasoning skills by generating questions directly from electronic health records. SCTonto is specifically designed for knowledge representation in SCT question generation, leveraging the depth and complexity of real-world medical data to create a more dynamic and relevant educational experience. This process enhances the SCT's applicability in medical education by providing students with practical, scenario-based questions that reflect the uncertainties inherent in clinical decision-making. The platform employs direct and indirect strategies for generating tests based on ontology-driven dataset.

In a similar vein Bratsas et al. [30] developed an ontology-based framework for generating medical computational problems within electronic learning environments, using ontologies to semantically enrich and interlink educational content and healthcare standards. This approach aims to streamline the creation and dissemination of medical educational materials, enhancing their interoperability and semantic coherence. This paper highlights the effectiveness of ontologies in organizing and presenting medical educational content in a manner that improves accessibility and learning outcomes, thereby advancing the pedagogical methods employed in medical education through enhanced semantic integration.

2.4.2 Simulation-Based Learning Systems and Applications in Medical Education

The evolution toward simulation-based education in medicine is widespread in the realm of medical educational applications. So, H. Y., et al. [31] investigated various simulators and came to the conclusion that simulations provide a structured environment where learners can refine clinical skills, navigate mistakes, and assimilate knowledge without the potential for harming patients. The use of simulation addresses several educational needs: controlling the sequence of tasks for learners, providing opportunities for guidance, preventing unsafe situations, and allowing the practice of rare or complex scenarios. Such environments can bridge the gap between theory and practice, ensuring that learners are equipped with the necessary skills in a controlled setting.

The adoption of virtual patients and immersive learning environments represents a significant part of medical simulators. These tools are effective for their ability to closely replicate clinical scenarios, offering interactive experiences that promote clinical reasoning and decision-making in a no-risk framework (Danforth et al. [32]).

The diversity of simulation tools, ranging from low to high-fidelity simulators and part-task trainers to comprehensive computer-based systems, is extensively discussed in the academic literature. High-fidelity simulations and virtual reality platforms are particularly lauded for their effectiveness in simulating real-life medical scenarios, thereby enhancing the learning experience by offering practical, hands-on engagement that traditional learning modalities cannot (Al-Elq [33]; Gayef [34]). These tools are indispensable for facilitating a seamless transition of theoretical knowledge to clinical application, providing dynamic and authentic learning settings. Despite its advantages, challenges such as technological limitations and cost are acknowledged.

Pai et al. [35] draws the same conclusions and also underscores the critical need for the integration of simulation-based medical education into the core medical curriculum as a strategy to enhance its effectiveness and ensure its sustainability. The authors argue for the indispensable role of faculty training and commitment, elucidating that the success of simulated-based programs extends beyond the mere provision of technical and equipment support to include the active involvement of educators proficient in simulation pedagogy.

2.4.3 Optimizing Medical Education with HCI

The domain of HCI is instrumental in refining the user experience (UX) and interface design, with its principles benefiting educational technologies.

Rundo et al. [36] evaluate HCI's impact on clinical decision-making processes, emphasizing its capacity to support physicians' cognitive tasks in precision medicine. This review underscores HCI's potential in medical education, where it can streamline complex data management and decision-making workflows, thereby improving training outcomes for healthcare professionals. The paper highlights the importance of HCI in creating intuitive user interfaces for medical software and systems, which are crucial for enhancing the usability and effectiveness of clinical tools. Furthermore, although this paper does not specifically address automated exercise generation, it contains many valuable insights applicable to such developments. The application of this thesis project imitates the real medical activities reviewed in this paper, drawing on the principles of HCI to simulate authentic clinical decision-making scenarios. This approach ensures that even in the realm of automated exercise generation, the system can provide a realistic and beneficial educational experience for healthcare professionals, closely mirroring the challenges and complexities of actual medical practice.

2.5 Existing Educational Applications for Medical Students that Simulate the Diagnosing Process

In the landscape of medical education, numerous educational applications have emerged to assist medical students and professionals in enhancing their diagnostic skills. These tools not only facilitate the learning process but also integrate technological features that mimic real-world clinical scenarios, thus providing an immersive learning experience.

"Isabel Healthcare" [37] stands out as a diagnostic aid tool designed to enhance clinical reasoning. It utilizes a comprehensive database of symptoms and conditions to generate differential diagnoses, effectively acting as a second opinion. This system supports medical students by expanding their diagnostic capabilities and improving decision-making through exposure to a wide array of potential conditions. The main feature of "Isabel" is its ability to integrate seamlessly with electronic health records (EHRs).

"VisualDx" [38] is another significant player in the field, known for its extensive library of medical images. This application aids in the diagnosis of dermatological conditions but also covers a broad spectrum of internal medicine. "VisualDx" is particularly valuable in medical education for its ability to visually demonstrate the variation in presentations of diseases, thus helping students and clinicians improve their diagnostic accuracy and speed. Its unique feature, the "VisualDx" differential diagnosis builder, enhances clinical education by allowing users to compare visually similar conditions side-by-side.

"DXplain" [39], developed by Massachusetts General Hospital, is a decision support system that provides a list of potential diagnoses based on the clinical attributes entered by the user. This tool serves a dual educational role, offering both a learning resource for medical students and a practical tool for experienced clinicians needing to refresh their differential diagnosis skills.

"InSimu Patient" [40] offers a unique approach by allowing users to interact with virtual patients. This application provides an environment where medical students can practice their clinical skills without risk to real patients. Each case scenario is designed to mimic the complexity of real-life diagnostics, complete with feedback and a scoring system to track progress and learning outcomes. "InSimu"'s standout feature is its algorithm that simulates real-life diagnostic paths, allowing students to experience the consequences of their clinical decisions.

"Prognosis: Your Diagnosis" [41] engages users with clinical case scenarios across various specialties, challenging them to apply their knowledge in diagnosing and managing diseases. Each case is crafted to reflect real-world medical situations, making it a useful tool for active learning and assessment. The application enhances learning through immediate feedback on decisions and a detailed explanation of the best diagnostic and treatment pathways for each case.

"Case" [42] provides medical students and professionals with daily clinical cases through a subscription model. This app is used for exam preparation and continuous medical education, as it offers detailed explanations and discussions on each case. It offers a personalized learning experience, adjusting the difficulty of cases based on the user's performance and learning pace.

2.6 Existing Research Gaps

Studies on ontology often focus on generating tests but rarely on content creation in simulations. While ontologies are extensively used for creating structured knowledge bases and generating test items (Leo et al. [28], Radović et al. [29]), there is a gap in research concerning the application of ontologies for developing content in simulation-based learning systems. Current literature primarily addresses the theoretical frameworks and methodologies for test generation, leaving the potential of ontologies in enhancing the realism and educational value of simulation scenarios underexplored.

Studying the role of feedback in learning applications in the medical domain is rare. The review of existing automated feedback systems (Gütl et al. [25], Demaidi et al. [26]) reveals an issue: the feedback provided is frequently generic and lacks specificity. This is particularly problematic in medical education, where detailed, context-specific feedback is crucial for effective learning. Not many systems align feedback with the actual clinical activities and decision-making processes students engage in, thereby diminishing the practical utility of the feedback provided.

There is a lack of research that studies ontology-driven content generation in simulation-based learning systems with ontology-driven feedback specifically in medicine. The integration of ontologies to create dynamic, contextually rich simulation scenarios and provide tailored feedback remains insufficiently explored. Existing studies focus either on the theoretical aspects of ontology application or on separate elements of simulation and feedback. Comprehensive research that combines ontology-driven content generation with personalized, ontology-based feedback in simulation-based learning environments is scarce. Addressing this gap is essential for developing more effective educational tools that can mimic real-life clinical situations and provide meaningful, actionable feedback to learners.

3. Design Process

3.1 Gathering Requirements

The first phase involves gathering materials and consulting with stakeholders, including MLX. The main goal of this stage is to understand the requirements and expectations, ensuring that the prototype developed aligns with the needs of medical students. A number of unstructured interviews were conducted with the CEO of MLX to gather necessary information about the Ear, Nose and Throat (ENT) domain and requirements for the application. Here are the main insights:

3.1.1 The Diagnosing Process of Ear-Related Diseases

MLX provided a detailed overview of the real-life diagnosing process for ear-related diseases (Figure 1). This information is needed for developing a realistic and comprehensive simulation.

Symptoms

The diagnostic process begins when a patient presents with symptoms. Patients usually report one to three symptoms, not revealing all the possible symptoms of their diagnosis. This step involves initial patient interaction, where the symptoms are documented.

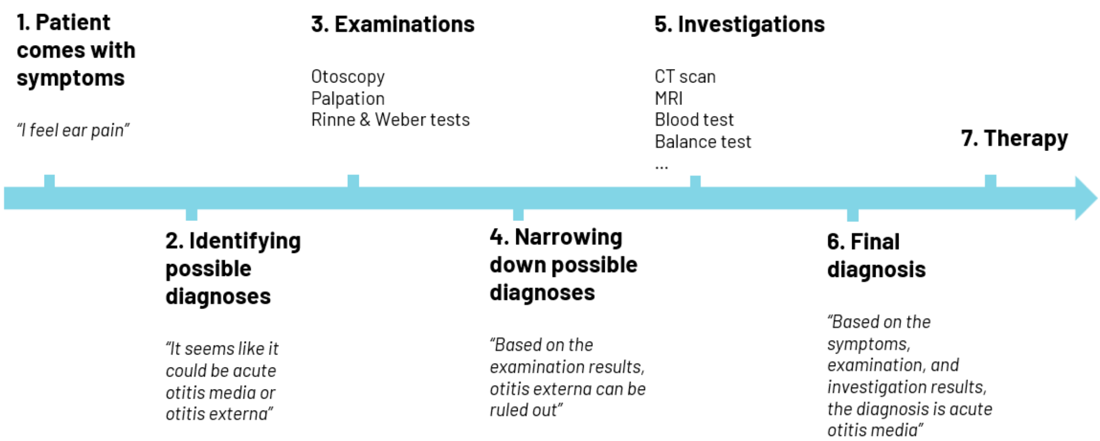

Figure 1. The scheme of the real-life diagnosing process for ENT disease.

Identifying Differential Diagnoses

Based on the presented symptoms, potential diagnoses are identified. The list of differential diagnoses helps doctors to narrow down the number of possible final diagnoses. For example, a symptom like tinnitus could suggest conditions such as otosclerosis or cholesteatoma.

Examinations

The next step involves conducting various examinations to gather more information. Common examinations for ear-related issues include otoscopy, palpation, and Rinne and Weber tests. These examinations can be conducted in the doctor's office during the first visit. They help refine the list of potential diagnoses by providing more specific data about the patient's condition.

Narrowing Down Differential Diagnoses

Based on the results of the examinations, the list of potential diagnoses is narrowed down. For instance, if otosclerosis is ruled out based on examination results, the focus can shift to other possible conditions. This step is necessary to ensure only essential investigations are conducted in the next phase.

Investigations

Further investigations, such as CT scans, MRIs, blood tests, and balance tests, are conducted to gather additional data. These procedures are typically conducted in specialized laboratories. These tests provide more detailed insights and are used as examination results to confirm or rule out specific conditions.

Final Diagnosis

After examinations and investigations, a final diagnosis is made. At this stage, the doctor has all the necessary information, including symptoms, examination results, and investigation results, to correctly identify the final diagnosis.

Therapy

At the final stage, the goal is to determine the appropriate therapy based on the diagnosis. In this thesis, we do not focus on this stage, so it can be reflected in the application superficially.

The objective of this interview is to incorporate these steps into the application so that the process of diagnosing is as realistic as possible.

3.1.2 Requirements for the Application

The application should be part of the MLX learning process, allowing students to train and check their knowledge while studying ENT diseases. The content for the exercise generation in this application should consist of the content of these materials. The materials were provided as 23 PDF and Excel files with semi-structured information, including: diagnoses, their background, symptoms, differential diagnoses, necessary examinations that should be conducted and their results, investigations and their results, and therapy to treat the diagnoses. The application should mirror the real process of diagnosing as discussed above. Therefore, every stage of the process should be reflected in the application. The generated exercise should be presented as a patient's story about their condition, including three symptoms of the selected diagnosis. Feedback should be provided for each user's correct, partially correct, and incorrect answers.

3.1.3 Insights for Ontology Creation and Expansion

Many insights were gathered for more effective creation of ontology, such as relationships between all the entities of the diagnosing process.

- Symptoms have three types of severity: mild, moderate, and severe.
- Symptoms can be sorted into 12 main groups based on the nature of the symptom: dizziness, ear discharge, hearing loss, ear pain, ear fullness, facial and cranial symptoms, general symptoms, inflammation, nausea and vomiting, tinnitus, sensory perception, and skin-related symptoms.
- Many diagnoses can share the same symptoms.
- Every diagnosis has at least five differential diagnoses. Only five differential diagnoses for each diagnosis were provided in the given materials. Some differential diagnoses are unique to particular diagnoses, while some are common among others.
- Some differential diagnoses are among the 23 given diagnoses, so one of the 23 targeted diagnoses can be a differential for another.
- Diagnoses can be split into two groups: 1) diagnoses of the inner ear and 2) diagnoses of the outer and middle ear. Based on this separation, particular examinations and investigations can be conducted.

- There are four examination techniques for ear disease: otoscopy, palpation, Rinne test, and Weber test.
- Otoscopy provides a visual examination of ears and ear-related locations, revealing hearing loss, ear fullness, ear discharge, inflammation, and skin-related symptoms.
- Palpation examines if there is pain in the ear area. In many cases, otoscopy and palpation should be conducted to see normal results. Abnormal results of otoscopy and palpation can be perceived as additional symptoms.
- Some examination results are unique to particular diagnoses, while some are common for many diagnoses.
- Some diagnoses don't require all four examination techniques. In some cases, palpation or Rinne and Weber tests are redundant.
- For every diagnosis, a doctor needs to conduct all necessary examinations.
- There are many investigation techniques that can be conducted, sorted into five categories: Radiology, Function tests, Laboratory tests, Microbiology, and Pathology, with 18 actual investigation techniques in total.
- Not every investigation is needed to identify the final diagnosis. Moreover, in real life, it is necessary to be cost-effective and conduct only the necessary investigations to identify the final diagnosis. This parameter is out of the scope of this prototype.
- The majority of investigation results are unique to each diagnosis; some can be shared by several diagnoses. Some diagnoses don't need investigations at all.
- There are four types of treatment: medical treatment, surgical treatment, wait and see, and wait and scan. Each diagnosis can have one to all four possible therapies.
- Therapy depends on the severity of symptoms and is unique for the majority of diagnoses. Medical treatment and surgical treatment have many types of treatments, while wait and see and wait and scan are as they are.
- Therapy is the final stage in the diagnosing process but not the most important in this project and less detailed. The main focus is on identifying

the final diagnosis based on the given information: symptoms, examination and investigation results, and differential diagnoses.

3.1.4 Nuances of Real-Life Doctor-Patient Interaction

In the provided materials, symptoms were described in medical language, which is very specific and not typically used by patients. For a more realistic interaction in the application, every symptom should be transformed from specific medical language to common terms used by real patients. For example, tinnitus should be described as persistent ringing in the head, and otalgia should be described as ear pain.

4. Implementation

4.1 Ontology Design

In this phase, the ontology is developed and populated using the materials provided during the requirement gathering stage.

The ontology was designed using the Web Ontology Language for knowledge representation. The content of the ontology was implemented based on the insights from the interviews with the MLX CEO. Classes, object properties, data properties, and individuals were created in the ontology.

Protégé [43] software was used to design and populate the ontology. Protégé is a widely used ontology editor that allows for the creation and management of ontological structures. Using Protégé, the relationships between different entities were defined, enabling the ontology to support dynamic and contextually relevant content generation within the application. The ontology was populated with the data from the MLX materials that contain information about 23 otology diagnoses.

4.1.1 Classes

All given 23 diagnoses are individuals of the class Diagnosis. Class Diagnosis has SensorineuralDiagnosis and ConductiveDiagnosis subclasses for diagnoses with sensorineural hearing loss and conductive hearing loss, respectively. Having these subclasses makes it possible to provide feedback about mistakes in refining differential diagnoses based on Rinne and Weber tests results. Additionally, there is a DifferentialDiagnosis subclass for diagnoses that are possible alternatives to the actual diagnosis, partially sharing mutual symptoms.

The Symptom class allows the arrangement of patient symptoms. This class has the subclasses, which are useful to split all symptoms into 12 groups. These subclass names reflect the names of symptoms: Dizziness, EarDischarge, HearingLoss, EarPain, EarFullness, FacialAndCranialSymptom, GeneralSymptom, In-

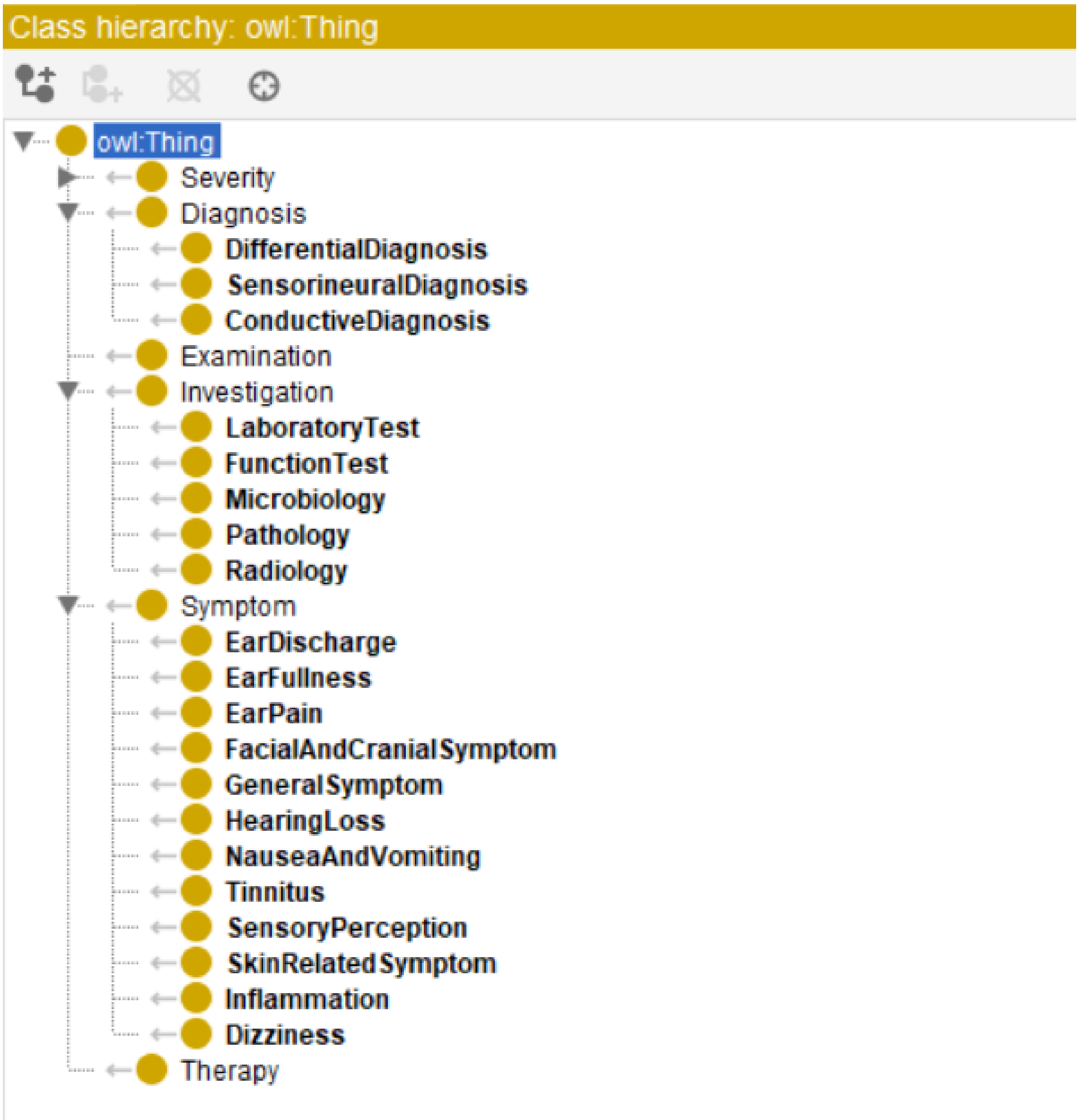

Figure 2. The class structure of the ontology, presented in the Protégé software.

flammation, NauseaAndVomiting, Tinnitus, SensoryPerception, and SkinRelatedSymptom.

The Examination class was created to manage clinical examinations such as otoscopy, palpation, Rinne, and Weber tests. Similarly, the Investigation class is used to manage investigations. Investigations are split into five main groups with the same names for subclasses: Radiology, FunctionTest, LaboratoryTest, Microbiology, and Pathology.

The Therapy class models treatment for the patient after the final diagnosis was revealed.

The structure of the classes can be found in Figure 2.

4.1.2 Object Properties

There are four object properties in this ontology that describe the relations between entities.

hasDifferentialDiagnosis relation is of the Diagnosis class and links diagnoses with all their differential diagnoses, which also belong to the Diagnosis class.

hasSymptom object property links diagnoses of the Diagnosis class with all relevant symptoms of the Symptom class.

leadsToDiagnosis relation is between examination results of the Examination class and relevant subclasses and diagnoses of the Diagnosis class.

needsTherapy property connects diagnoses of the Diagnosis class with appropriate treatment options of the Therapy class.

4.1.3 Labels and Data Properties

Labels in annotations reflect the names of diagnoses, symptoms, examination and investigation techniques, and therapy. Additionally, there is a group of data properties called hasExaminationResults and hasInvestigationResults which serve to contain the text of examination and investigation results for diagnoses. All data properties are texts and have strings as a range for datatype.

4.1.4 Individuals

All 23 diagnoses were added as individuals with the class Diagnosis, one of the subclasses SensorineuralDiagnosis or ConductiveDiagnosis, the label diagnosisName, and the object properties hasSymptoms, hasDifferentialDiagnosis, and needsTherapy (Figure 3).

Individuals of examinations have distinctive names for each diagnosis. These names consist of the name of the diagnosis and the examination technique, for example, Otosclerosis-otoscopy, Otosclerosis-palpation. These individuals have the data property hasExaminationResults with the text of otoscopy, palpation, Rinne, and Weber tests results. Also, these individuals have the object property leadsToDiagnosis to connect examination results with a particular diagnosis. Totally 84 individuals of examination results were added to the ontology.

In a similar way, individuals for investigations were implemented with the Investigation class and one of the five subclasses of investigation techniques. Their names contain the diagnosis name, the investigation group, and the investigation technique itself, for example, Otosclerosis-Radiology-CT, Otosclerosis-Function-tests-Audiometry. These individuals also have the label for the name of the investigation technique. As with examinations, investigation individuals have the leadsToDiagnosis object property to maintain the relation between the diagnosis and the investigation results. 98 individuals of investigation results were incorporated into the ontology

Symptoms individuals belong to the Symptom class and one of twelve subclasses, which reflect the group of symptoms they are in. Symptoms have the label with the name of the symptom written in simple language that patients usually use. 134 different symptoms were added to the ontology.

Individuals for therapy belong to the Therapy class and have the label with the name of the therapy in it. There are only four kinds of therapy names: medical treatment, surgical treatment, wait and see, wait and scan.

There are also 50 diagnoses that belong only to the DifferentialDiagnosis subclass of the Diagnosis class, with the label of their names and the hasSymptom object property. These diagnoses are unique differential diagnoses for several main diagnoses, so there is no information about them in the given MLX materials, except common symptoms with the main diagnosis and their names.

The schematic representation of the final ontology entities and relations can be found in Figure 4.

4.2 Application Design

Based on the interviews with the CEO of MLX, the application should be a web application so that it can be integrated with digital theoretical learning materials. The backend, built with Python, handles ontology interactions, while the frontend is built with HTML for structure, CSS for styling, and JavaScript for dynamic behavior and interactivity.

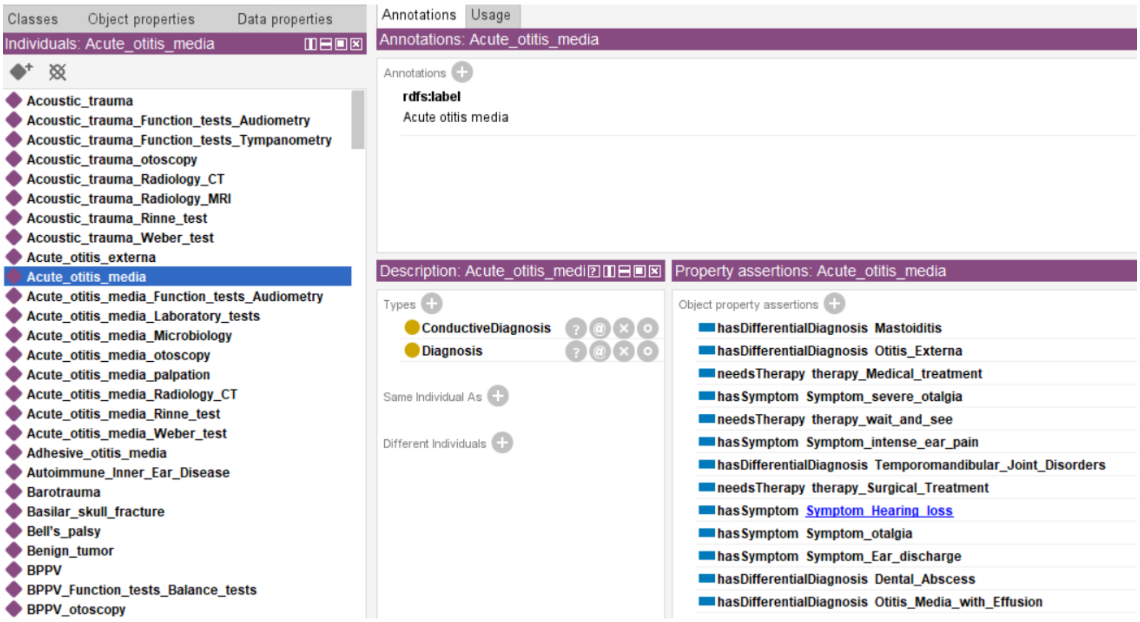

Figure 3. The ontology individuals, presented in the Protégé software.

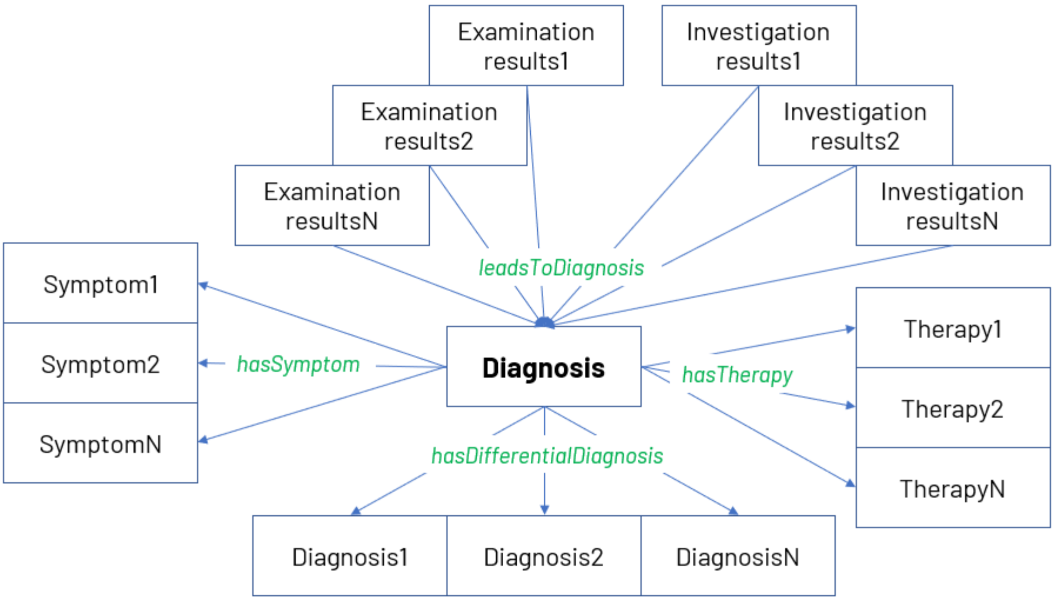

Figure 4. A schematic representation of ontology elements and their relationships with each other.

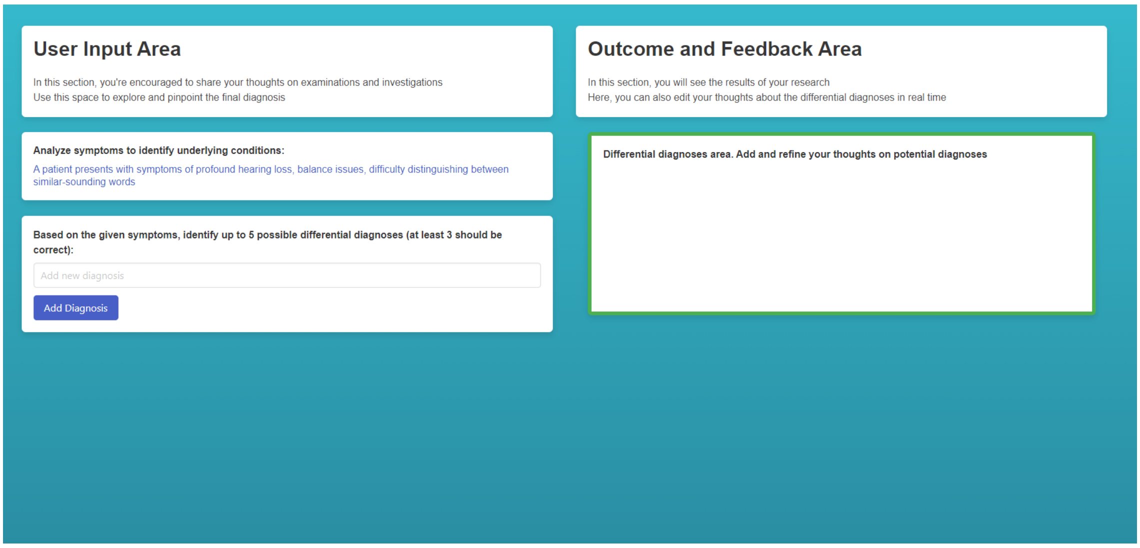

Figure 5. The start page of the first prototype features a two-column interface. The generated exercise highlighted in blue font.

4.2.1 First Prototype

The interface of the first prototype was designed with two columns (Figure 5).

- First Column

This column generates the exercise with the patient's symptoms, selecting three random symptoms of a randomly selected diagnosis from the 23 main diagnoses (Figure 5).

Input fields are provided where the user can input their answers to the main questions:

What are the differential diagnoses for the given symptoms? The next fields appear only if the user correctly identifies at least three out of six possible diagnoses. The maximum number of differential diagnoses is six, including the final diagnosis and its five differential diagnoses. The "Submit Diagnoses" button appears only if the user has selected four diagnoses. Clicking this button submits the answer (Figure 6). If the user did not guess at least three differential diagnoses correctly, the session breaks and the "Start Over" button appears, clicking on which the new session with the new generated exercise started.

What examinations should be conducted for the given symptoms? This input field appears after successfully submitting the differential diagnoses. To add an examination technique and obtain the examination results, the user needs to type the name of the technique and click the "Submit Examination" button. To

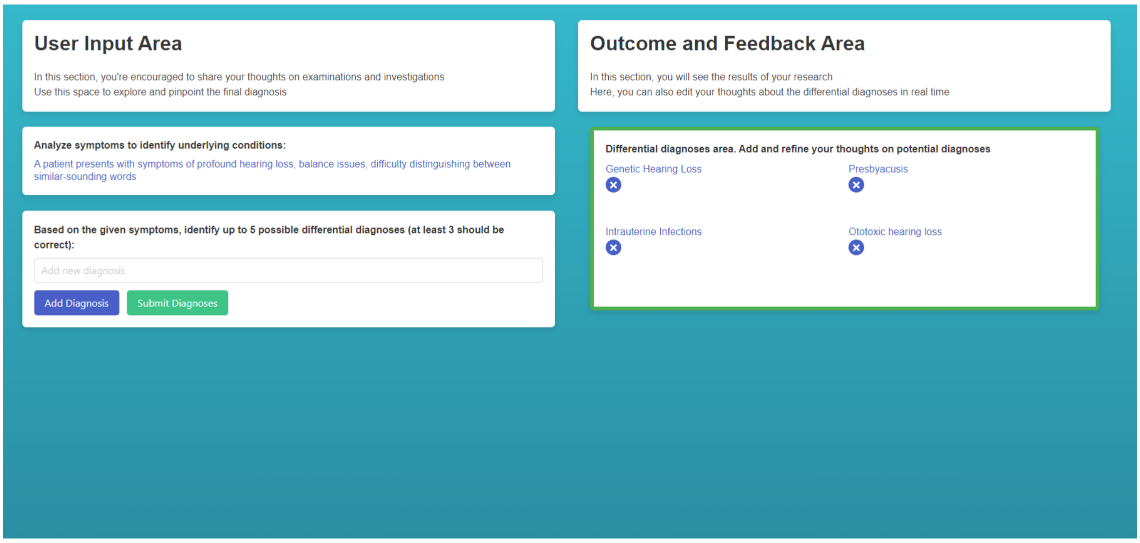

Figure 6. Four differential diagnoses are added via the input field. The "Submit Diagnoses" button appears after three diagnoses have been added.

finish this stage, the "Finish Examinations" button should be clicked (Figure 7).

What investigations should be conducted for the given symptoms? This field appears after the examination stage is finished. The input here consists of five buttons, each representing a group of investigation techniques. By clicking on these buttons, smaller buttons with the actual investigation techniques appear. Clicking on these smaller buttons, the user enters the necessary investigations to obtain the investigation results. To finish this stage, the user needs to click on the "Finish Investigation" button (Figure 8).

What is the final diagnosis? The input field for the final diagnosis appears after clicking the "Finish Investigations" button. The user needs to type the name of the final diagnosis and click the "Submit Final Diagnosis" button (Figure 9).

What is the appropriate therapy for the final diagnosis? The therapy input field appears if the final diagnosis is identified correctly. This input field is implemented as a drop-down menu with four types of possible treatment methods (Figure 10). To submit the answer, the user needs to click the "Submit Therapy" button (Figure 11).

- Second Column

This column displays the results of the user's answers and feedback so the user has all the necessary information for identifying the final diagnosis. The main fields here are:

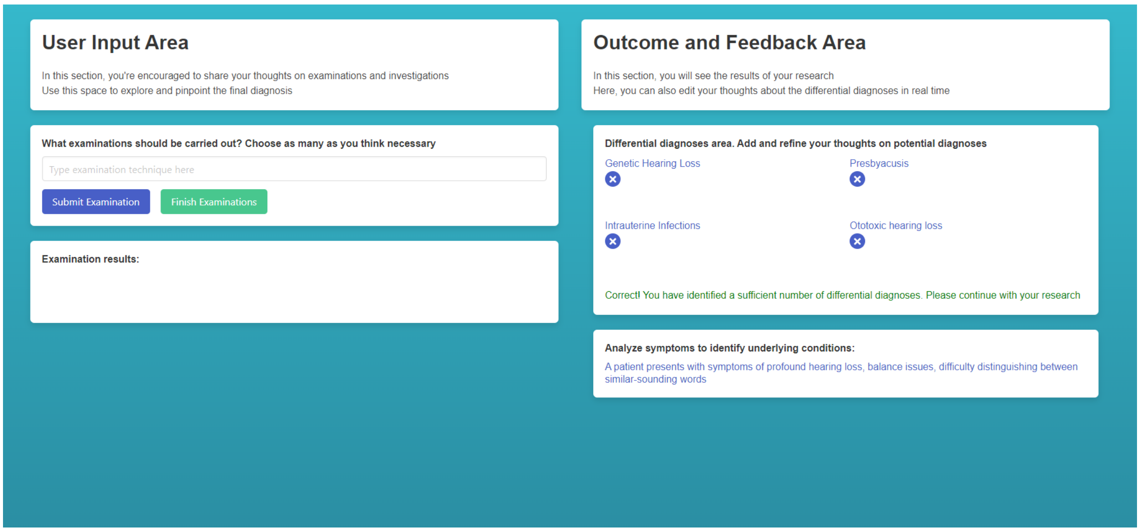

Figure 7. The examination input field, "Submit Examination", and "Finish Examinations" buttons. The list of differential diagnoses selected by the user and the positive feedback on the user's answer.

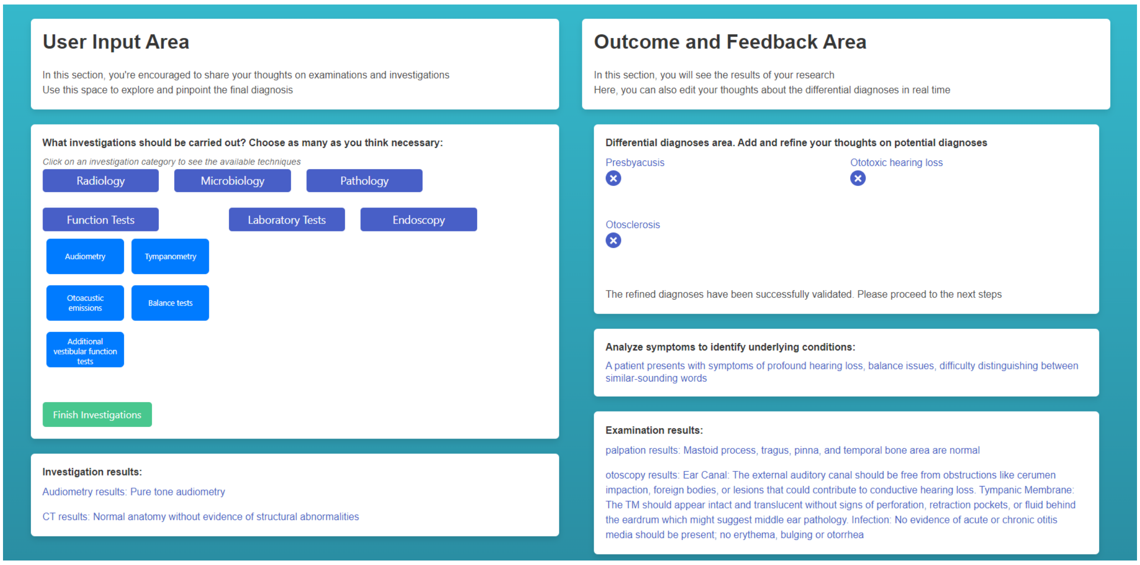

Figure 8. The investigation techniques groups, actual investigation techniques, and "Finish Investigations" buttons. The examination results field.

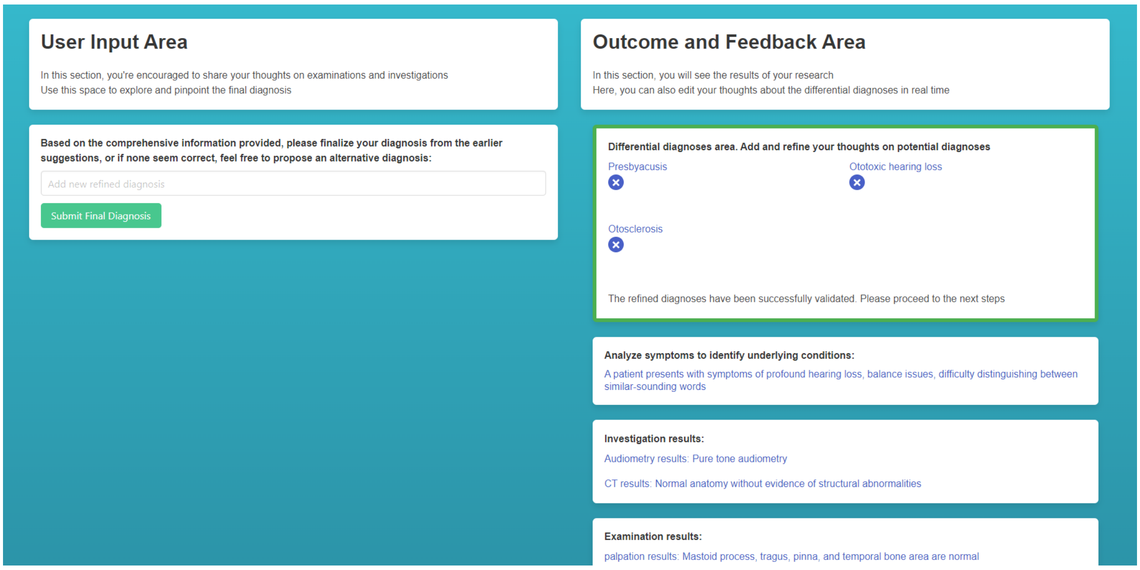

Figure 9. The final diagnosis input field and the "Submit Final Diagnosis" button. The investigation results field.

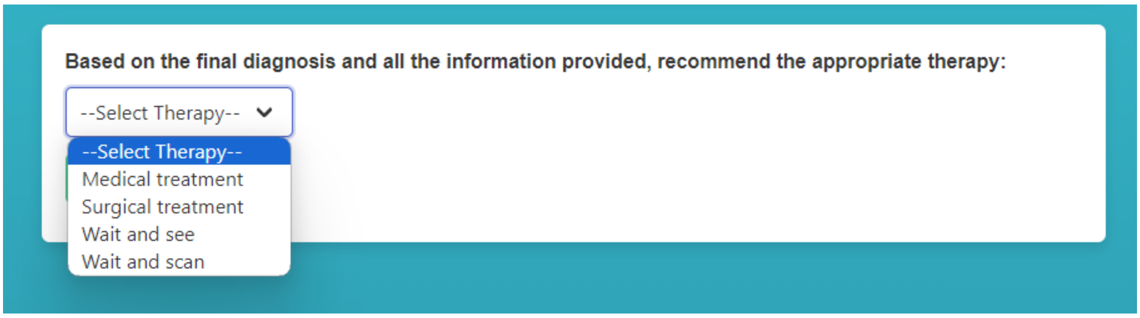

Figure 10. The therapy options drop-down menu.

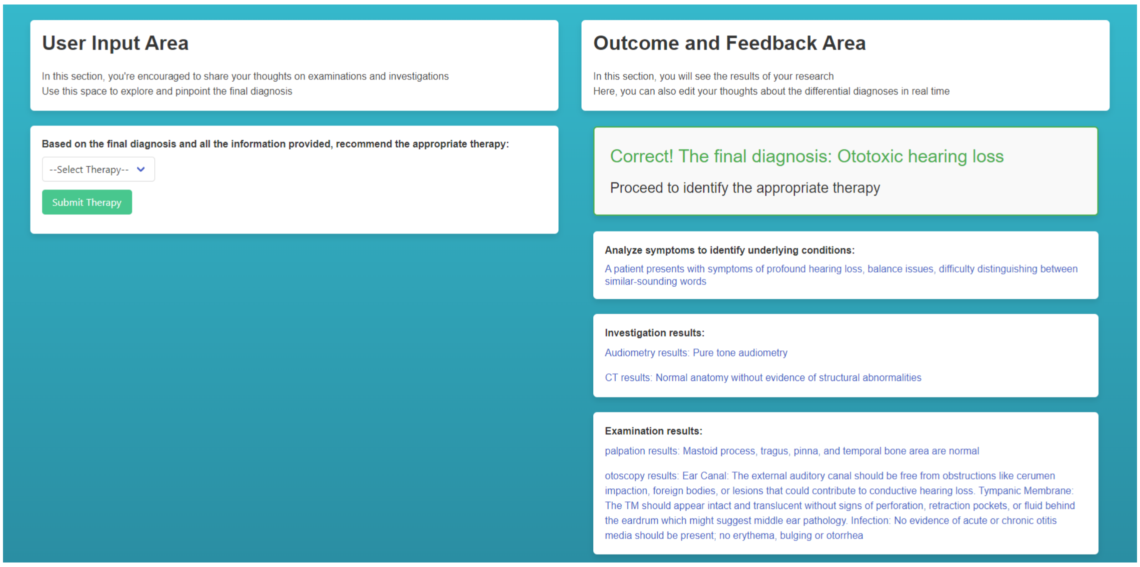

Figure 11. The therapy input and final diagnosis feedback fields.

Differential Diagnoses. The user can see and adjust possible diagnoses based on the given symptoms. They can delete (using a "x" button under the selected differential diagnosis name) and add a new diagnosis before submitting their answer. Feedback appears at the bottom of this field after clicking the "Submit Diagnoses" button in the input differential diagnoses area. The feedback can be positive or negative (Figure 7).

Examination Results. The results of the selected examination techniques are displayed here after clicking on the "Finish Examinations" button (Figure 8).

Investigation Results. Similar to the examination results, but for investigation techniques (Figure 7).

Final Diagnosis and Therapy Feedback. The field with feedback on the correctness of the final diagnosis submission appears instead of the differential diagnoses list field (Figure 11). The same applies to the therapy stage.

In this prototype version, the user must type the full names of differential diagnoses, examination techniques, final diagnoses, and therapies in the input fields. The input for investigations is made through five buttons representing the five groups of investigation techniques. When a user clicks on these buttons, smaller buttons appear with the names of specific investigation techniques.

If the user does not select the differential diagnoses correctly, the round stops, and a "Start Over" button appears that refreshes the page and generates a new exercise. To get a new exercise, the page must be refreshed.

4.3 First Prototype Testing

The testing of the first prototype was conducted with six medical students who were learning otolaryngology at MLX. They were instructed about the purpose of the study, and brief instructions about the application's functionality were provided. The session lasted 30 minutes. The students were in one room with the experiment supervisor and used the application on iPad Pro devices provided by MLX.

4.3.1 First Prototype Testing Insights

The feedback on the first prototype was gathered through unstructured interviews with the participants. The main insight was about the difficulty in correctly typing the names of three differential diagnoses. The system required an exact match of diagnosis names, and none of the students managed to progress past the first stage. Consequently, feedback for the other parts of the application's functionality was not gathered.

Two new requirements were suggested to adjust the application based on the testing results:

- Implement autocomplete for search suggestions in the text input fields, so the user can select the appropriate answer based on the first letters they type. This will relax the requirement for exact matching of diagnosis names in the ontology.
- Remove the session termination rule after incorrect completion of the identifying differential diagnoses stage. Instead, apply the termination rule only for the final diagnosis stage.

Additional requirements were suggested during an interview with the CEO of MLX after the first prototype testing:

- Introduce an additional stage for refining the differential diagnoses list after gathering the examination results. At this stage, the user can delete differential diagnoses entered during the first stage and add new ones. The refined list should consist of three diagnoses in the end.
- Modify the current two-column interface to display comprehensive data of

the patient's symptoms, the list of differential diagnoses, and the examination and investigation results without scrolling. This change is necessary to allow users to see all the information at once, facilitating a more accurate identification of the final diagnosis.

- The generated exercise should be more realistic, so instead of just giving the user three symptoms, the exercise should appear as a patient's story describing their conditions.

4.4 Second Prototype

The second prototype was developed based on the feedback gathered from the first prototype testing. The interface design of the second prototype includes a static input field located at the bottom of the screen, similar to those in messengers, chats, and search engines, making it familiar and intuitive for users. The area where feedback on the user's actions is displayed is now fixed in one place next to the input field, ensuring it is always visible. The main part of the screen now serves as the area where the generated exercise, the list of differential diagnoses, and the results of examinations and investigations are displayed. This layout allows users to see all necessary information without the need for scrolling. Additionally, new user interface elements were implemented, such as info buttons for each stage to elaborate on the task, navigation to highlight the current stage, and a brief tutorial on how to navigate the interface, which can be accessed at any time by clicking the "Tutorial" button. The purpose of these elements is to make the application more user-friendly and easy to use. Since the second prototype was designed to test the full application functionality, an ontology-driven feedback system was developed to provide users with comprehensive information on their actions.

4.4.1 Second Prototype Interface

The second prototype user interface is divided into four main parts (Figure 12):

Input Area: Located at the bottom of the interface, this area is designated for user responses. Users can click the Info button (Figure 13) at any stage for details about the current task. An autocomplete feature helps in suggesting possible answers based on user input, with exceptions during the investigation (Figure

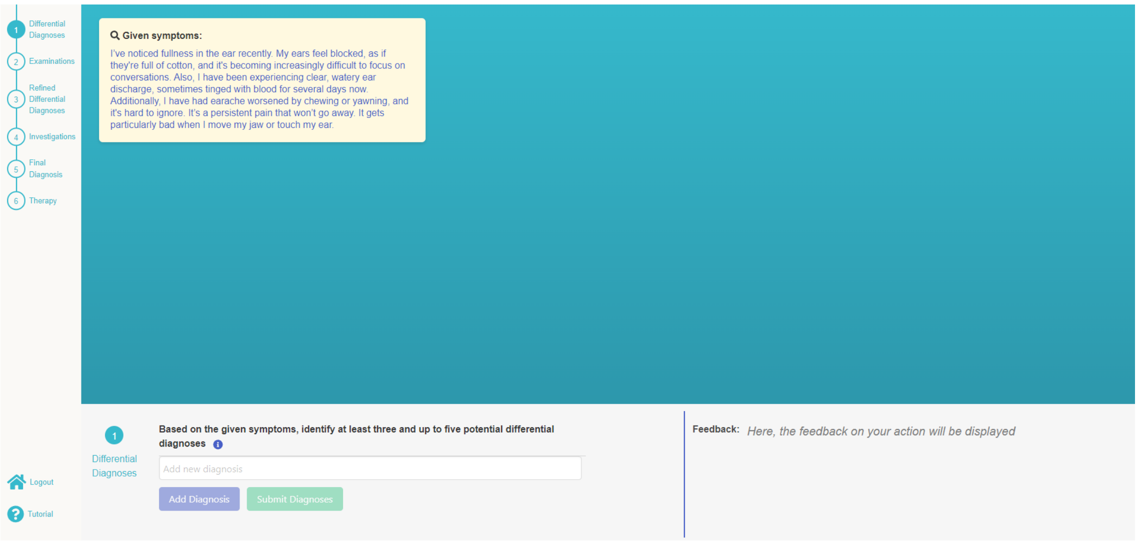

Figure 12. The start page of the second prototype. The navigation area, the input and feedback fields, and the working area with the generated exercise.

14) and therapy stages, where selections are made through buttons similar to the first prototype.

Feedback Area: To the right of the input area, this segment offers immediate feedback on user inputs using a color-coded system — green for correct, orange for partially correct (as depicted with an orange icon and orange font for the feedback text in Figure 15), and red for incorrect responses. After the submission of the results by clicking "Submit Diagnoses" button, the buttons with hints appear in the feedback field. At different stages, different hints could be provided (Figure 15). Progression to the next stage is made possible with the "Next" button that appears instead of the current input field.

Working Area: This is the central display area where the generated exercise is shown. Starting with a randomly generated symptoms, it guides the user through identifying potential diagnoses, conducting examinations to obtain results, refining the list of differential diagnoses, conducting further investigations to gather more information about the final diagnosis, and finally selecting the appropriate therapy. This process is depicted across various stages, with elements appearing sequentially at every stage (Figure 16), ending with the typing of the final diagnosis with all necessary information at hand.

After successfully identifying the final diagnosis and therapy, a "Start Over" button appears (Figure 11). By clicking on this button, users can initiate a new exercise, beginning the process anew with a different generated exercise.

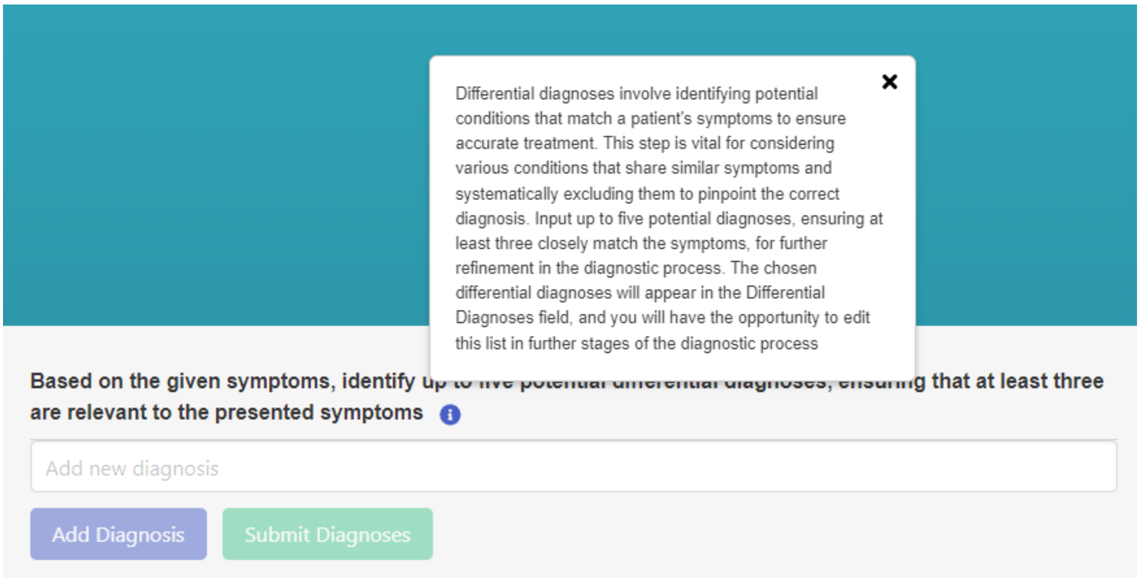

Figure 13. Text bubble showing detailed information about the current stage's task via the "Info" button.

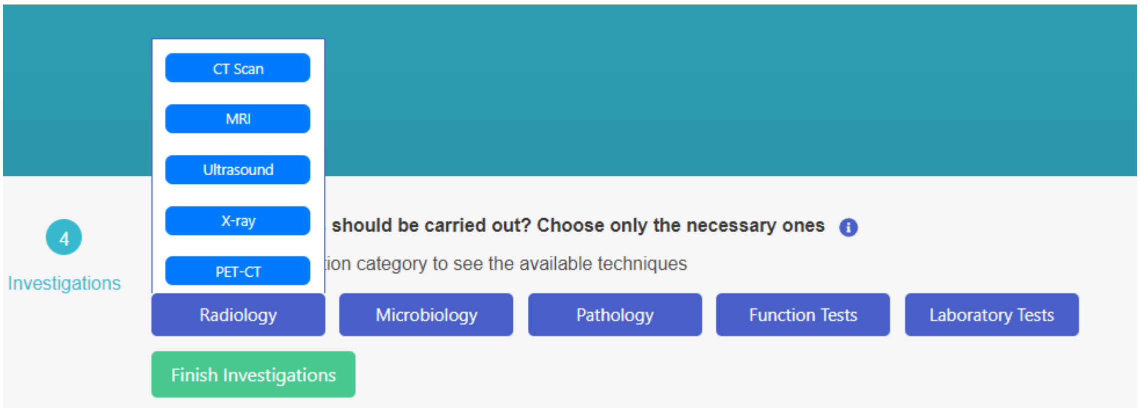

Figure 14. Investigations stage input area.

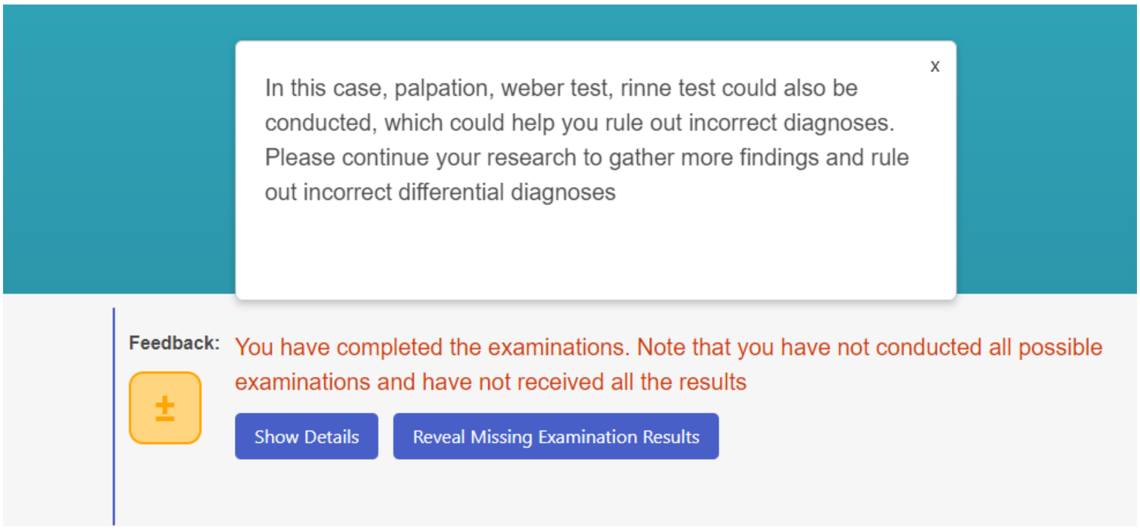

Figure 15. The hint buttons at the examination stage and the semi-correct answer hint text.

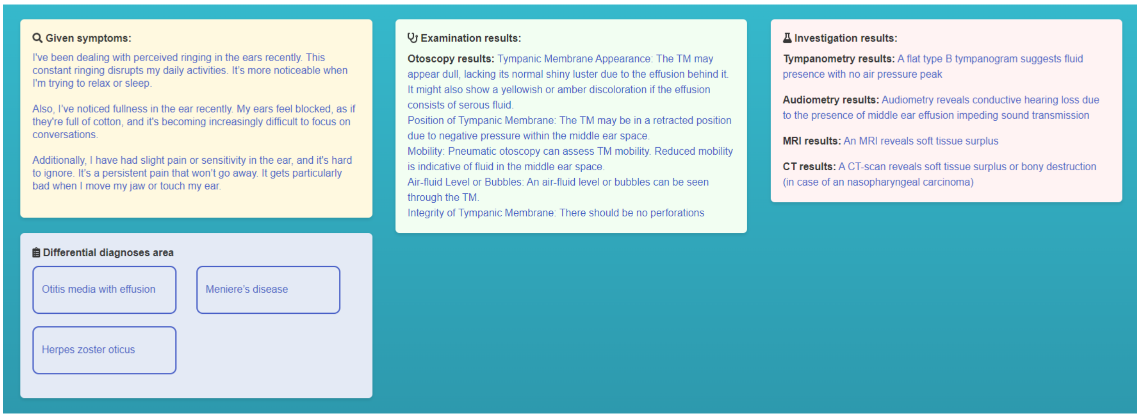

Figure 16. All working area elements displayed.

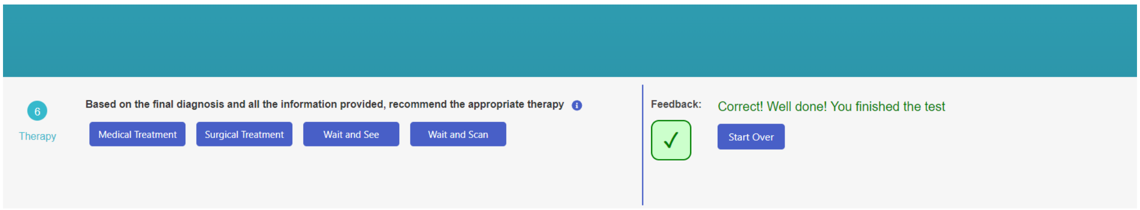

Figure 17. Final feedback and Start Over button.

Navigation Area: Shows the user's progress through a progress bar and includes a "Tutorial" button (Figure 12).

After completing the exercise and successfully identifying the final diagnosis and therapy, a "Start Over" button appears (Figure 17). By clicking on this button, users can initiate a new exercise, beginning the process anew with a different case.

4.4.2 Second Prototype Ontology-driven Feedback

The feedback system in the second prototype was enhanced with ontology capabilities to provide comprehensive feedback on possible incorrect or partially correct answers from users. This feedback system was designed based on the work of Hattie and Timperley[21]. In this study, the feedback parameters suitable for computer-based educational systems were utilized. According to their findings, effective feedback should:

- Point out specific mistakes related to the task.
- Provide actionable steps for improvement in subsequent steps.
- Be immediate for tasks that are part of the whole exercise.
- Acknowledge correct answers.

The feedback in the second prototype operates on three levels:

- First Level - Simple Feedback.

This level provides basic feedback about the correctness of the task. A short message indicates whether the answer is correct, partially correct or incorrect. Visual cues include green font for correct answers, orange for partially correct answers, and red for incorrect answers. Distinctive symbols accompany the feedback: green square with the "V" symbol for correct answers, orange square with the "±" symbol for partially correct answers, and red square with the "X" symbol for incorrect answers. Correct answers are acknowledged with an encouraging message and a prompt to proceed to the next step with no additional hints.

- Second Level - Detailed Feedback (First Hint).

For partially correct and incorrect answers, the first hint elaborates on the initial feedback message. This hint can be accessed by clicking the "Show Details" button and appears in a bubble that points out the mistake and explains why it is a mistake.

- Third Level - Reveal Answer Feedback (Second Hint).

For partially correct answers, a second hint is also provided. This hint, accessible via a button in the feedback field, reveals one incorrect answer and describes why it is incorrect. For incorrect answers, the second hint works similarly but reveals one correct answer instead. Due to the limited information provided by MLX materials, it was challenging to implement detailed messages explaining why the revealed correct answers are accurate. Therefore, the second hint for incorrect answers is not as detailed as desired.

Identifying Differential Diagnoses Stage.

During the process of identifying the correct differential diagnoses, the user should select at least three and up to five diagnoses. If all the diagnoses selected by the user are correct, including the final diagnosis, the correct answer feedback appears in the feedback field with the short message: "Excellent! All differential diagnoses you have selected are correct. Further examination will help you gather more information about the patient."

If the user selects at least one correct differential diagnosis or the final diagnosis itself, partially correct feedback appears in the feedback field, indicating that some of the selected differential diagnoses are incorrect. By clicking on the first hint "Show Details" button, the system checks in the ontology which symptom groups of the three generated patient's symptoms in the exercise do not align with the incorrect differential diagnoses' symptom groups. This shows that the incorrect differential diagnoses' symptoms do not match the given symptoms. In the feedback bubble, there will be a list of messages indicating that one or more selected differential diagnoses do not align with the particular given symptoms. If more than one incorrect differential diagnosis does not align with the same given symptoms, the message transforms into "Some of the differential diagnoses do not align with ..." to avoid counting similar messages in the feedback bubble.

In rare cases when the incorrect differential diagnosis aligns with the given patient's symptoms but is still incorrect due to a distinctive symptom that defines this diagnosis as not suitable, the message in the bubble will be "One or more of these diagnoses align with the given symptoms, but have a distinctive symptom that makes this differential diagnosis incorrect." Due to limited information in the provided materials, it is impossible to manage this feedback in detail. The second hint, presented as the "Reveal One Incorrect Diagnosis" button, indicates the incorrect refined differential diagnosis by marking it in red in the differential diagnoses list. Additionally, a feedback bubble appears, displaying the details that do not match for this specific diagnosis.

The same first hint feedback logic also applies to incorrect answers. In most cases, the first hint bubble message for incorrect answers will indicate that the selected differential diagnoses do not align with all the given patient's symptoms. The second hint, provided by the "Reveal One Correct Diagnosis" button, will disclose one correct differential diagnosis by highlighting it in green in the differential diagnoses area, but it won't include a detailed explanation because the provided materials lack sufficient information.

Examinations Stage.

The examinations stage is when the user gathers additional information about the patient's condition. The examination results can be perceived as additional patient symptoms that help the user narrow down possible diagnoses. Consequently, there are no incorrect answers at this stage, and therefore, no incorrect answer feedback.

If the user selects all necessary examination techniques, the correct answer simple feedback appears in the feedback field with the message: "Excellent! You have completed the necessary examinations. Continue by refining the list of possible diagnoses."

Partially correct answers can vary. If the user does not select all necessary examination techniques, the partially correct simple feedback message will be: "You have not conducted all necessary examinations." By clicking on the "Show Details" button, the message in the first hint bubble will inform the user about which examination techniques could also be conducted. Additionally, if the user misses otoscopy or palpation, the ontology can identify if the given patient's symptoms require otoscopy for visual examination (e.g., ear discharge, ear fullness, inflammation, skin-related symptoms) or palpation for examining ear pain. The message in the feedback bubble can inform the user that otoscopy or palpation is necessary due to the presence of specific symptoms. The second hint, that can be provided by clicking on the "Reveal Missed Investigation Results" button, reveals random missed examination results in the examination results field and marks them in red.

Another potential mistake is selecting redundant examinations. If the user selects an examination technique that is not necessary for the given symptoms, the simple feedback message will be: "You have conducted a redundant examination." In the first hint bubble, the message will indicate that the redundant examination techniques yielded no useful results. The second hint will not appear if the other necessary examinations were selected correctly.

Finally, a combination of mistakes can be addressed in feedback messages. If the user selects a redundant examination technique and does not select all necessary ones, the feedback will reflect both errors.

Refining Differential Diagnoses Stage.

The refining differential diagnoses stage is similar to the identifying differential diagnoses stage but includes additional information to consider—examination results. At this stage, the user needs to narrow down the list of possible diagnoses to three.

If all three refined differential diagnoses are correct, the simple feedback message will be: "Excellent! All refined differential diagnoses you have selected are correct. Further investigations will help you gather more information about the patient."

If at least one refined differential diagnosis is correct while others are not, the first feedback message will be: "Some of the refined differential diagnoses you selected are incorrect," and two hint buttons will appear. By clicking the "Show Details" button, the first hint bubble will pop up with three possible combinations of messages:

- Mismatch with Symptoms:

Similar to the message from the identifying differential diagnoses stage, it shows that some differential diagnoses do not align with the particular patient's symptoms.

- Mismatch with Rinne and Weber Test Results:

The ontology can identify if any of the incorrectly refined differential diagnoses align with the Rinne test or Weber test results. If these tests show that the final diagnosis has conductive or sensorineural hearing loss while the selected differential diagnosis has the opposite type (e.g., Rinne and Weber tests show conductive hearing loss and the user selected a differential diagnosis with sensorineural hearing loss), the message will be: "Some of the selected differential diagnoses seem to involve conductive hearing loss and don't align with the Rinne and Weber test results, which show sensorineural hearing loss."

- Mismatch with Otoscopy or Palpation Results:

This message shows the mismatch between otoscopy or palpation results, perceived as additional symptoms, and the symptom groups of the incorrect differential diagnoses. For example: "Some of the selected differential diagnoses do not align with the otoscopy results, which show ..." followed by the particular symptom group, such as ear discharge or inflammation. The second hint reveals the incorrect refined differential diagnosis by highlighting the diagnosis in red along with the feedback bubble showing the mismatched details for the particular diagnosis.

The simple feedback for an incorrect answer is: "Unfortunately, all of the selected refined differential diagnoses are incorrect." The first hint message will be similar to the partially correct answer, but in most cases, it will indicate that the selected diagnoses do not align with the patient's symptoms and examination results. The second hint will reveal one correct differential diagnosis in the same way as it was done in the identifying differential diagnoses stage.

Investigations Stage.

During the investigations stage, the user needs to select the necessary investigation techniques to gather more information about the patient and identify the final diagnosis. Due to the lack of information about specific investigation results from the provided materials, it is impossible to design additional relations in the ontology between investigation results and other entities. Therefore, the feedback at this stage is not ontology-driven. The ontology here is used as a source of correct and incorrect investigation techniques specific to the selected diagnosis.

Similar to the examinations stage, there are no incorrect answers at this stage. If the user selects all possible investigations, the simple feedback message will be: "Excellent! You have completed all possible investigations. Continue by identifying the final diagnosis."

The partially correct feedback is also similar to the examinations stage. If the user misses possible investigation techniques, the short feedback message will be: "You have not conducted all possible investigations." The first hint button, "Show Details," reveals a bubble with a message indicating which missed investigations could be conducted to gather more useful information about the patient. The second hint button, "Reveal Missed Investigation Results," reveals random missed investigation results in the investigation results field and marks them in red. If the user selects unnecessary investigation techniques, the simple feedback message will be: "You have not completed all possible investigations, yet some of the investigations you conducted were redundant." The first hint message will inform the user that particular investigation techniques yielded no useful results. The second hint will not appear if the other necessary investigations were selected.

If the user misses some investigations and simultaneously selects unnecessary ones, a mixed message of the previous options will appear for all kinds of feedback.

Identifying Final Diagnosis and Therapy Stage.

At the last two stages, identifying the final diagnosis and therapy, the feedback is simple and not ontology-driven. If the user selects the final diagnosis incorrectly, the feedback message will reveal the correct final diagnosis: "Unfortunately, that is not the right diagnosis. Try again. The final diagnosis is ...". If the user identifies the final diagnosis correctly, the feedback message will be: "Excellent! This is the right diagnosis. The final diagnosis is ...".

For identifying the appropriate therapy, if the user selects the correct therapy, the feedback message will be: "Excellent! You finished the test. Well done!" If the user selects the incorrect therapy, the feedback message will be: "Incorrect, try again."

5. Evaluation

The evaluation stage is designed to assess the effectiveness of the ontology-driven simulation-based learning system. This process involves a structured approach to gather quantitative and qualitative data from medical students using the application. The evaluation is carried out in several steps as detailed below.

5.1 Participants

The evaluation involves six medical students provided by MLX who are currently learning otology. They were recruited with the help of the CEO of the company. The only requirement was that participants be over 18 years old.

5.2 Evaluation Design

The evaluation design involves pre- and post-test knowledge evaluations of the students, with the application being used between the tests to measure the knowledge gained. The pre-test contains six questions aimed at assessing the initial knowledge of the students. The post-test's goal is to measure the knowledge gained after using the application. The post-test contains six isomorphic questions about the same subjects as in the pre-test. Both tests focused on three diagnoses selected by the CEO: acute otitis media, cholesteatoma, and vestibular schwannoma. These three diagnoses reflect the variety of procedures that students can perform and receive detailed feedback on before determining the final diagnosis. The decision to focus on only three diagnoses was due to the 35-minute time limit for the application evaluation stage. These tests were implemented using the Qualtrics survey design platform provided by Utrecht University.

Additionally, logs of user actions in the application will be gathered to analyze session durations, inputs, buttons clicked, and the correctness of answers.

Finally, a user experience evaluation questionnaire is designed to measure how engaging and helpful the application is, how supportive the interface is, how effective the feedback is, and the overall user experience.

The application for the evaluation session was modified so that the first three generated diagnoses are acute otitis media, cholesteatoma, and vestibular schwannoma, as the questions in the pre- and post-tests include only these diagnoses. After these three diagnoses, the generated exercises proceed randomly.

Ideally, to better evaluate the effectiveness of the ontology-driven feedback component of the application, the experiment should also include a testing group using the application with a simple, non-ontology-driven feedback system. However, due to the limited availability of medical students studying otology at MLX and their intensive educational schedules, it was decided to omit this part of the evaluation.

5.3 Procedure

The participants were placed in a room under supervision of the CEO of MLX. Each student was assigned a unique number (1-6) to be used as their login ID and password for the application. All participants used iPad Pro devices while filling in the surveys and testing the application.

The CEO of MLX sent the participants a link to a Qualtrics survey (can be found in Appendix A), which includes both the pre-test and post-test, as well as the link to the application and logins and passwords to access the application. The session duration is 60 minutes and is structured as follows:

5-8 minutes: The evaluation session begins with instructions on the purpose of the study, the timing, and an overview of the application interface elements.

5-7 minutes: Participants complete the pre-test via the provided survey link.

3 minutes: After completing the pre-test, participants select their assigned number from a menu on the next page. Based on their choice, the next page displays their unique login and password. The link to the application is also provided on this page, along with short instructions and a timer counting down from 35 minutes.

35 minutes: Participants click the link, log into the application, and begin testing it.

5-7 minutes: After testing the application, participants return to the survey page to complete the post-test.

While the participants use the application, their actions are logged. The logged actions include: user login, start of the session (generated exercise), the generated final diagnosis, the clicked buttons, typed words (letter by letter), submitted answers, and timestamps. This allows for the analysis of the correctness of users' answers at every stage, their behavior within the session, which user interface elements were used and which were not, the duration of sessions, and the time spent on each stage.

A week after the controlled session, the user experience questionnaire (can be found in Appendix B) was sent to the participants. During this week, the students had access to the application and were encouraged to use it. The application was modified so that exercises were randomized from the beginning, unlike the controlled session where it began with the three initial diagnoses reflected in the pre- and post-tests. The questionnaire contains questions about the user experience, interface, the application's effectiveness in enhancing knowledge, and the usefulness of feedback. It includes 12 Likert scale questions, one Net Promoter Score (NPS) question on overall experience, and two open-ended questions about what the participants liked most while using the application and what should be changed or improved.

6. Results

The results were exported from Qualtrics as an Excel file. Data were gathered from all six participants, except one who did not finish the post-test, resulting in pre- and post-test evaluations from only five participants. Additionally, one participant did not complete the user experience questionnaire, so those results also come from five participants. However, all six participants logged into the application, providing comprehensive data on their actions within the application.

6.1 Pre- and Post-Test Results

The pre- and post-test results, along with the correct identification of the final diagnosis using the application, are shown in Table 1. While two users showed an increase in correct answers in the post-test, two users showed no progression, and one user showed a relatively significant decrease in correct answers compared to the pre-test. This result suggests that there is no significant difference between the pre-test and post-test scores. To establish more robust results, additional participants must be involved in the experiment. The detailed pre- and post-test results can be found in Appendix C.

6.2 Logs Analysis Results

The analysis of the application logs provided several key insights into the participants' interactions and learning progress during the evaluation session (Figure

	Pre-test (correct answers)	Post-test (correct answers)	Difference
User1	1	2	1
User2	1	2	1
User3	1	1	0
User5	3	1	-2
User6	2	2	0

Table 1. Table of pre- and post-tests results.

user1	Button Click	Button ID: submitSecondRefinedDiagnosesButton, Button Text: Submit Final Diagnosis	2024-06-28 16:27:55
user1	Input Change	Input ID: newSecondRefinedDiagnosis, Input Value: Tym	2024-06-28 16:27:53
user1	Input Change	Input ID: newSecondRefinedDiagnosis, Input Value: T	2024-06-28 16:27:52
user1	Input Change	Input ID: newSecondRefinedDiagnosis, Input Value: Ty	2024-06-28 16:27:52
user1	Button Click	Button ID: No ID, Button Text: Next	2024-06-28 16:27:48
user1	Button Click	Button ID: No ID, Button Text: x	2024-06-28 16:27:47
user3	Button Click	Button ID: No ID, Button Text: Reveal One Missing Investigation Result	2024-06-28 16:27:30
user1	Button Click	Button ID: No ID, Button Text: Show Details	2024-06-28 16:27:30

Figure 18. The section of the user logs from the application that displays the user's login, action name, action description, and the date and time of each action.

18).

General Insights.

All participants completed at least two diagnosis tasks of exercises. Two participants completed three diagnosis tasks. Two participants did not finish the diagnosis task. One participant finished two diagnosis tasks and then refreshed the exercise five times, receiving diagnoses outside the scope of the pre- and post-test questions. The first task completion averaged 17 minutes, while the second and third task completion averaged 11 minutes, resulting in an overall average task completion time of 13 minutes. All participants clicked on all hint buttons for detailed feedback in every diagnosis task. Four participants checked the info button for additional information.

Identifying and Refining Differential Diagnoses Stages.

Two participants selected five out of five possible differential diagnoses to submit, two participants selected four, and two selected the necessary minimum of three differential diagnoses. All participants edited the list of differential diagnoses several times before submitting. Participants edited the list of differential diagnoses more frequently with cholesteatoma than with acute otitis media and vestibular schwannoma. All participants selected at least one differential diagnosis involving the opposite type of disease, such as conductive or sensorineural hearing loss, during one of the tasks. Four participants did this during two diagnostic tasks. All the participants who selected the final diagnosis from the possible ones during the identifying or refining differential diagnoses stage correctly defined the final diagnosis. None of the participants decreased the number of correctly identified differential diagnoses during the refining stage. They either improved it or left it the same as in the defining differential diagnoses stage. Four participants identified no correct differential diagnoses at least once.

Examinations Stage.

Five participants confused examination techniques with investigation techniques during the first diagnosis task. One participant confused a diagnosis with a symptom. One participant had difficulty distinguishing between Dutch and English examination technique names. None of the participants typed the examination technique names correctly in the first diagnosis task. After finishing the first diagnosis task, participants typed all examination technique names correctly. This was probably due to a code error: if a user typed something in the examination input field, deleted it, and clicked the backspace key, all options for examination technique names appeared, allowing users to simply select them. This error was detected by users according to the user experience questionnaire, which can be found in the Open-Ended Responses section below.

Investigations Stage.

All participants selected almost all possible investigation techniques in the first diagnosis task.

6.3 User Experience Questionnaire Results

The results of the user experience questionnaire provided useful insights, which are depicted in Figure 19 and 20. The detailed user experience questionnaire results can be found in Appendix D.

"This application was useful overall."

All five participants agreed that the application was useful overall.

"This application helped me with some knowledge gaps."

Four out of five participants agreed with this statement, while one participant neither agreed nor disagreed.

"This application was engaging."

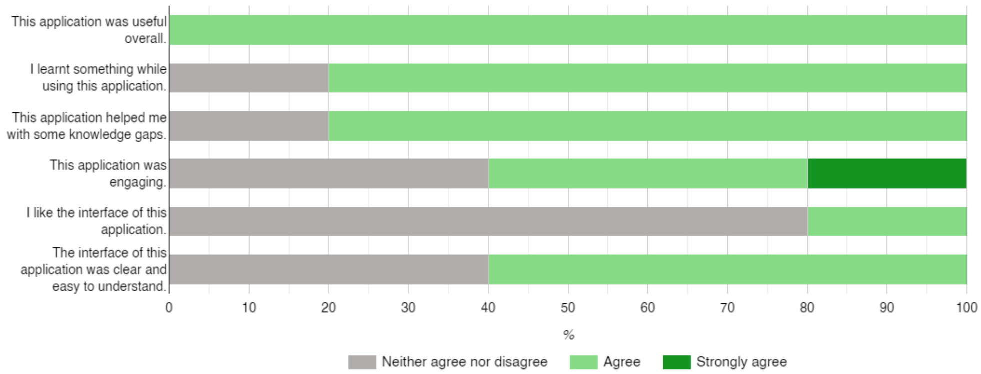

Figure 19. A stacked bar chart representing the distribution of Likert-scale statement agreements for the first six user experience questionnaire questions.

One participant strongly agreed, two participants agreed, and two participants neither agreed nor disagreed that the application was engaging.

"I like the interface of this application."

Only one participant agreed with this statement, while the others neither agreed nor disagreed.

"The interface of this application was clear and easy to understand."

Three participants agreed that the interface was clear and easy to understand, while two participants neither agreed nor disagreed.

"The feedback given by this application was useful." "The feedback given by this application was informative."

The same participants provided identical answers for these two questions. Three of them agreed that the feedback was useful and informative, while two others neither agreed nor disagreed.

"The feedback given by this application helped me progress through the exercises."

Four participants agreed with this statement, while one participant neither agreed nor disagreed.

"I feel that the exercises helped me learn Ear diagnoses."

One participant strongly agreed, three participants agreed, and one participant neither agreed nor disagreed.

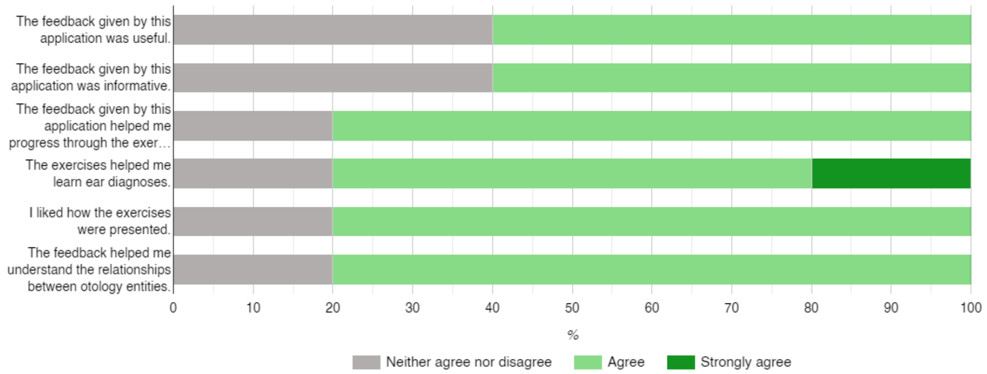

Figure 20. A stacked bar chart representing the distribution of Likert-scale statement agreements for the second six user experience questionnaire questions.

"I liked how the exercises were presented."

Four participants agreed with this statement, while one participant neither agreed nor disagreed.

"The feedback helped me understand the relationships between otology entities."

Four participants agreed with this statement, while one participant neither agreed nor disagreed.

"How would you rate your overall experience with the application?"

In response to this question, participants were asked to evaluate the overall user experience on a 10-point scale. Three participants rated the experience with scores of 8, 7, and 7, respectively, while two participants provided less favorable ratings of 6 and 6.

Open-Ended Responses

What did you like most about the application?

Participants provided the following responses (unedited for grammar):

"That it gives you feedback"

"Helping tool for making a plan"

"The fact that the cases were generated so that the student can learn continuously new things."

"The variety of cases, the feedback and hints were very helpful"

"Feedback buttons and how the exercise was presented"

What would you change or improve about the application?

Participants suggested the following improvements (unedited for grammar):

"More feedback and there are some loops in the application"

"Some bugs: when you refresh the page you get a new story"

"The options came all if you backspaced a couple of times. This decreased the learning effect. It would be more effective if the app could read what we all typed and afterwards asked if we mean by example 'otoscopy' in that way we will be forced to think first and afterwards corrected."

"More interactivity, pictures, more modern interface"

"Fix a bug, when erase examination name and press backspace once more, the options for examinations appear"

7. Discussion

7.1 Results Interpretation

Pre- and Post-Test Results

As there is no significant difference between the pre-test and post-test scores, no definitive conclusions can be drawn. Given the small sample size, the results may not be generalizable. More robust conclusions could be drawn with a larger participant pool to ensure statistical power and reliability.

Logs Analysis Results

The analysis of the application logs provided several key insights into the participants' interactions and learning progress during the evaluation session.

The average task completion time decreased from 17 minutes in the first task to 11 minutes in subsequent tasks, indicating a potential increase in efficiency or familiarity with the application. High usage of hint buttons for detailed feedback suggests that participants relied on these features for guidance, highlighting their importance in the learning process.

Initial confusion in identifying examination techniques and typing errors suggests a learning curve. The observed improvement in correctly typing examination techniques after the first diagnosis task indicates that participants adapted and learned over time, albeit with the aid of a detected code bug.

The variation in the number of differential diagnoses selected by participants (ranging from three to five) and the repeated editing of lists indicate an iterative learning process where users refined their choices as they gathered more information. Participants edited the list of differential diagnoses more frequently with cholesteatoma than with acute otitis media and vestibular schwannoma, suggesting that cholesteatoma presented more diagnostic challenges. All participants selected at least one differential diagnosis involving the opposite type of disease (such as conductive or sensorineural hearing loss) during one of the tasks, indicating a recurring error in understanding. This underscores the value of feedback addressing this mistake to help medical students learn this particular aspect. Participants who selected the final diagnosis from the possible ones during the identifying or refining differential diagnoses stage correctly defined the final diagnosis, demonstrating the integrity of diagnostic findings such as symptoms, examination, and investigation results reflected in the system. Four participants identified no correct differential diagnoses at least once, indicating that there were instances where the participants struggled with the complexity of the task. Notably, two of these four students correctly identified the final diagnosis in the end. This suggests that the hint revealing one correct answer helped half of this group of participants correctly identify the final diagnosis. None of the participants decreased the number of correctly identified differential diagnoses during the refining stage, indicating a positive progression in diagnostic accuracy.

The broad selection of investigation techniques in the first diagnosis task suggests initial uncertainty or a strategy of thorough testing. The absence of strict rules to limit the selection of necessary techniques, as in the real-world diagnostic process, may also explain the selection of almost all investigation techniques.

This analysis allowed to identify common mistakes that students make when diagnosing ear diseases and understand their knowledge gaps, which can be addressed in future research.

User Experience Questionnaire Results

The user experience questionnaire results provided additional insights into participants' perceptions of the application:

All participants agreed that the application was useful overall, indicating its value in enhancing their learning experience. Four out of five participants agreed that the application helped with some knowledge gaps, suggesting effectiveness in addressing educational needs for the majority of users. The feedback on engagement was mixed, with only one participant strongly agreeing and two agreeing that the application was engaging, while two neither agreed nor disagreed. This indicates potential areas for enhancing the application's ability to capture and maintain user interest. Additionally, while three participants found the interface clear and easy to understand, only one participant explicitly liked the interface. This points to a need for improvements in design and usability to make the interface more appealing and intuitive for all users. Participants generally found the feedback useful and informative. This underscores the importance of provided feedback in supporting learning processes. Additionally, the feedback helped the participants progress through the exercises, highlighting its role in aiding users' navigation through tasks. The exercises were seen as helpful in learning ear diagnostics. This indicates the exercises' presentation is effective. Similarly, four participants agreed that the feedback helped them understand relationships between otology entities, emphasizing the value of feedback. When asked to rate their overall experience, participants gave mixed ratings. This suggests a generally positive experience but highlights the need for improvements to achieve higher satisfaction ratings. The answers to open-ended questions support the results gathered from this questionnaire.

7.2 Contribution

This thesis makes several contributions to the design of training tools for medical students.

Through the pilot study, the research identified common mistakes and knowledge gaps among medical students in diagnosing ear diseases. This analysis is useful for understanding where students typically struggle and how the application can be tailored to address these deficiencies effectively in future system designs.

The use of ontology-driven feedback was shown to be valuable in guiding students through the diagnostic process. Participants frequently relied on detailed feedback, which aided their progression through exercises and improved their diagnostic accuracy over time. This indicates that the strategy for providing feedback was effectively defined.

The user experience questionnaire results highlighted the strengths and weaknesses of the current interface. While the application was generally found to be useful and informative, areas for improvement were identified, such as enhancing engagement and usability. These insights are useful for future interface development, ensuring that the application becomes more intuitive, engaging, and up to date.

The pilot study provided preliminary data on the application's effectiveness, indicating areas where further research is needed. The study's findings, though not statistically significant due to the small sample size, suggest potential benefits of ontology-driven feedback in medical education. The research sets the groundwork for more extensive studies with larger participant pools to validate these findings and further refine the application.

7.3 Limitations

Firstly, the highly specific nature of otology required extensive time and additional meetings with the CEO of MLX to fully understand the nuances of otology entities and their interrelations. This was necessary to accurately implement the logic into the ontology, exercise generation, and detailed feedback system, which added complexity and duration to the project.

Secondly, the limitations of time caused challenges in developing a high-quality, user-friendly, and engaging user interface. Creating an interface that meets the standards of applications in 2024 proved to be difficult, impacting the overall user experience and potentially the effectiveness of the application.

Finally, a significant limitation of this study was the difficulty in accessing medical students to conduct evaluations. The curriculum at MLX allocates only two weeks for learning ENT diseases, with even less time dedicated to ear diseases. This short timeframe made it challenging to gather students, administer the application, and have them complete pre- and post-tests as well as the user experience questionnaire. After this period, students were no longer available for participation and gathering feedback.

7.4 Future Work

First, due to the limitations of time and the difficulty in accessing medical students, no study has been conducted to compare the results of students who used the application with and without ontology-driven feedback. Such a comparison is necessary to more accurately determine the effectiveness of ontology-driven feedback. Therefore, it is planned to conduct future studies in sessions using two types of applications: one with ontology-driven feedback and one without, involving two different groups of medical students. This will allow for a controlled comparison to isolate the impact of the ontology-driven feedback on learning outcomes. According to agreements with the CEO of MLX, approximately 60 participants will be provided by the end of 2024 to evaluate the application. It is vital for the study to have a notable number of participants to gather statistically significant results.

Additionally, the gaps in the relationships of investigation results and the possible feedback that could be provided at this stage will be addressed. This includes expanding the significance of the treatment identification stage based on symptoms, achieved through the addition of more comprehensive theoretical materials. Additionally, the DifferentialDiagnosis subclass of the Diagnosis class and the Severity class were created for future work and are not used in the logic of the current ontology and application. Moreover, adding new classes and subclasses is being considered. For example, implementing a NonOtologyDiagnosis class for differential diagnoses unrelated to ear diseases would create additional useful for feedback generation relationships between the ontology entities. Enhancements will also be made to the feedback mechanisms to provide more detailed and context-specific guidance, thereby improving the learning experience.

There are also future plans to expand the application and ontology capabilities to cover other ENT diseases. This expansion will involve developing new modules and exercises for diseases related to the nose and throat, thus broadening the scope and utility of the application for medical students.

Finally, there are plans to write a scientific paper in collaboration with Dr. Sergey Sosnovsky and the CEO of MLX, Jochen Bretschneider. This paper will detail the methodology, findings, and implications of the research, and it will be submitted to a peer-reviewed journal. After that, a presentation of these findings is planned for the International Educational Conference in 2025.

8. Conclusion

This thesis aimed to design and evaluate a prototype application that generates medical case studies and provides detailed ontology-driven feedback to medical students, with a focus on diagnosing ear diseases. The contributions of this work are manifold, offering insights into system design, the practical application of ontology in medical education, and the user experience of such a system. The pilot study demonstrated that ontology-driven feedback could be valuable in guiding students through the diagnostic process. Despite the small sample size, the frequent use of detailed feedback by participants and their progression through exercises suggest that this approach is effective in improving diagnostic accuracy and supporting the learning process. However, the lack of significant improvement between pre- and post-test scores highlights the need for a larger participant pool to validate these preliminary findings. Through detailed log analysis, the study identified common mistakes and knowledge gaps among medical students. These insights are critical for understanding where students struggle and how the educational tool can be tailored to address these deficiencies. The iterative learning process observed, with participants refining their choices and improving their accuracy, underscores the potential of the application to enhance medical training. The user experience questionnaire provided valuable feedback on the strengths and weaknesses of the current interface. While the application was generally found to be useful and informative, the need for improvements in design and usability was evident. Enhancing engagement and making the interface more intuitive and appealing are essential steps for future development. The study faced several limitations, including the highly specific nature of otology, time constraints in developing a user-friendly interface, and difficulties in accessing a sufficient number of medical students for evaluation. These limitations impacted the overall findings and highlight areas for future work. Future studies should include comparisons between applications with and without ontology-driven feedback to more accurately determine its effectiveness. Additionally, expanding the application's capabilities to cover other ENT diseases will broaden its utility for medical students. Collaborations for scientific publications and presentations at international conferences are also in the pipeline.

Overall, this thesis has laid the groundwork for the integration of ontology-driven feedback in medical education applications. While preliminary results are promising, further research with a larger sample size is necessary to validate these findings and refine the application. The insights gained from this study provide a solid foundation for future development and research, with the ultimate goal of enhancing the training and education of medical students.

Appendices

A. Appendix

Pre- and post-tests survey

Informed consent

Welcome and thank you for your participation in this study!

The purpose of this study is to evaluate the effectiveness of an application designed to assist in learning the diagnostic process in otology. Using the application, you will identify the final diagnosis by conducting differential diagnoses, as well as performing examinations and investigations.

Study Procedure:

Pre-Test: You will first complete a 7-minute pre-test to assess your initial knowledge in diagnosing otology conditions.

Application Testing: After the pre-test, you will use the application to identify differential diagnoses, conduct necessary examinations and investigations, determine the final diagnosis, and suggest appropriate therapy. This part of the study will take about 35 minutes.

Post-Test: Finally, you will complete a 7-minute post-test similar to the pre-test. This test will help us understand how much you have learned.

Participation in this study is entirely voluntary. You may withdraw at any time without any consequences. If you decide to withdraw, any data collected up until that point will be anonymized and retained for the research.

We hope you enjoy working with the application and kindly ask you to put in your best effort and approach the process thoughtfully. Good luck!

This study is being conducted by Andrei Marinin as part of a master's thesis in Human Computer Interaction at Utrecht University, under the supervision of Dr. Sergey Sosnovsky. This study is anonymous, and no personal data will be collected. You will be assigned a non-personal unique user ID and a unique password to use the application and complete the tests. All data gathered will be anonymized and securely stored. If you have any questions or concerns about this research, please contact Andrei Marinin at a.a.marinin@students.uu.nl.

○ I confirm that I have read and understood the above statements, and agree to participate in the study

○ I do not consent, I do not wish to participate

Pre-test

Which symptoms are associated with acute otitis media? (select all that apply)

☐ Hearing loss

- [ ] Facial palsy
- [ ] Swelling behind the ear
- [ ] Intense ear pain
- [ ] Ear discharge
- [ ] I don't know

Patient shows facial numbness, hearing impairment, balance issues, but no pain. What could be the diagnosis? (select all that apply)

- [ ] Mastoiditis
- [ ] Herpes zoster oticus
- [ ] Vestibular schwannoma
- [ ] Acute otitis media
- [ ] Presbyacusis
- [ ] I don’t know

Which of the following are potential differential diagnoses for cholesteatoma? (select all that apply)

- [ ] Acute otitis media
- [ ] Adhesive otitis media
- [ ] Lateral skull base fractures
- [ ] Exostoses and osteomas
- [ ] Dental abscess
- [ ] I don’t know

What do Weber and Rinne tests results demonstrate in acute otitis media? (select all that apply)

- ☐ Positive Rinne test in the affected ear
- ☐ Negative Rinne test in the affected ear
- ☐ Weber test lateralizes to the affected ear
- ☐ No lateralization on Weber test
- ☐ Rinne test positive on both sides
- ☐ I don’t know

What would otoscopy **NOT** reveal in a patient with vestibular schwannoma? (select all that apply)

- ☐ Normal tympanic membrane
- ☐ Fluid behind the tympanic membrane
- ☐ No signs of tympanic membrane inflammation
- ☐ No signs of tympanic membrane perforation
- ☐ Perforated tympanic membrane
- ☐ I don’t know

What investigations are required for diagnosing cholesteatoma? (select all that apply)

- ☐ No additional investigations needed
- ☐ CT scan
- ☐ MRI
- ☐ X-ray
- ☐ Ultrasound
- ☐ I don’t know

Block 4

Please select the user number that was assigned to you at the beginning of the study. Based on your selection, you will be provided with a login and password on the next screen for application testing.

Note: If you do not remember your user number, please ask the study moderator in the room.

Using the application

You have finished the pre-test. Now it is time to use the application. **Note!** Do not close this window until you have completed the post-test.

1. **Open the application** in a new tab using this link: https://quizitor.science.uu.nl/mlx-eval/login

2. Fill in the login and password:
Login: user1
Password: otosclerosis

3. Read the short instruction that will be displayed after logging in

4. **Use the application** for 35 minutes

5. After 35 minutes, **come back to this page**. The button to proceed to the next page will appear after 35 minutes

6. **Click the yellow button** with a black arrow at the bottom of this page to proceed to the post-test

You have finished the pre-test. Now it is time to use the application. **Note!** Do not close this window until you have completed the post-test.

1. **Open the application** in a new tab using this link: https://quizitor.science.uu.nl/mlx-eval/login

2. Fill in the login and password:
Login: user2
Password: cholesteatoma

3. Read the short instruction that will be displayed after logging in

4. **Use the application** for 35 minutes

5. After 35 minutes, **come back to this page**. The button to proceed to the next page will appear after 35 minutes

3500

Post-test

You have tested the application. Please now fill in the post-test. It will take 7 minutes.

Which of the following are potential differential diagnoses for cholesteatoma? (select all that apply)

☐ Acute otitis media
☐ Chronic Otitis Media with Effusion
☐ Lateral skull base fractures
☐ Herpes zoster oticus
☐ Squamous cell carcinoma
☐ I don’t know

Patient shows ear discharge, hearing loss, and intense ear pain. What could be the diagnosis? (select all that apply)

☐ Acute otitis media

- [ ] BPPV
- [ ] Vestibular schwannoma
- [ ] Dental abscess
- [ ] Inflammation of the auricle
- [ ] I don't know

Which symptoms are associated with vestibular schwannoma (select all that apply)

- [ ] Hearing loss
- [ ] Facial numbness
- [ ] Ear pain
- [ ] Ear discharge
- [ ] Balance issues
- [ ] I don’t know

What investigations are **NOT** required for diagnosing cholesteatoma? (select all that apply)

- [ ] Audiometry
- [ ] CT scan
- [ ] MRI
- [ ] X-ray
- [ ] Ultrasound
- [ ] I don’t know

What would otoscopy reveal in a patient with vestibular schwannoma? (select all that apply)

- ☐ Normal tympanic membrane
- ☐ Fluid behind the tympanic membrane
- ☐ Signs of obstruction or infection in the external auditory canal
- ☐ Granulation tissue is presented in the ear canal
- ☐ Perforated tympanic membrane
- ☐ I don’t know

What do Weber and Rinne tests results **NOT** show in acute otitis media? (select all that apply)

- ☐ Positive Rinne test in the affected ear
- ☐ Negative Rinne test in the affected ear
- ☐ Weber test lateralizes to the affected ear
- ☐ No lateralization on Weber test
- ☐ Rinne test positive on both sides
- ☐ I don’t know

B. Appendix

User experience questionnaire

Product Satisfaction

Welcome and thank you for your participation in this study!

You have recently used the application designed to assist in learning the diagnostic process in otology. Thank you once again for participating in this study! The purpose of this survey is to gather your feedback on various aspects of the application. Your responses will help us understand your experience and improve the application's design and functionality. This questionnaire will take 5 minutes of your time.

Participation in this study is entirely voluntary. You may withdraw at any time without any consequences. If you decide to withdraw, any data collected up until that point will be anonymized and retained for the research.

This study is being conducted by Andrei Marinin as part of a master's thesis in Human Computer Interaction at Utrecht University, under the supervision of Dr. Sergey Sosnovsky. This study is anonymous, and no personal data will be collected. All data gathered will be anonymized and securely stored. If you have any questions or concerns about this research, please contact Andrei Marinin at a.a.marinin@students.uu.nl.

- ◯ I confirm that I have read and understood the above statements, and agree to participate in the study
- ◯ I do not consent, I do not wish to participate

This application was useful overall.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

I learnt something while using this application.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

This application helped me with some knowledge gaps.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree

- ○ Disagree
- ○ Strongly disagree

This application was engaging.

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree

I like the interface of this application.

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree

The interface of this application was clear and easy to understand.

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree

The feedback given by this application was useful.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

The feedback given by this application was informative.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

The feedback given by this application helped me progress through the exercises.

- ◯ Strongly agree
- ◯ Agree
- ◯ Neither agree nor disagree
- ◯ Disagree
- ◯ Strongly disagree

I feel that the exercises helped me learn Ear diagnoses.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

I liked how the exercises were presented.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

The feedback helped me understand the relationships between otology entities.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

How would you rate your overall experience with the application?

Very bad | | | | | | | | | | Very good
0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○

What do you like most about the application?

What would you change or improve about the application?

C. Appendix

Pre- and post-tests results

Pre-test results

User	Which symptoms are associated with acute otitis media? (select all that apply)	Patient shows facial numbness, hearing impairment, balance issues, but no pain. What could be the diagnosis? (select all that apply)	Which of the following are potential differential diagnoses for cholesteatoma? (select all that apply)	What do Weber and Rinne tests results demonstrate in acute otitis media? (select all that apply)	What would otoscopy NOT reveal in a patient with vestibular schwannoma? (select all that apply)	What investigations are required for diagnosing cholesteatoma? (select all that apply)
user6	Hearing loss,Intense ear pain,Ear discharge	Herpes zoster oticus,Presbyacusis	Adhesive otitis media,Exostoses and osteomas	Positive Rinne test in the affected ear,No lateralization on Weber test	No signs of tympanic membrane inflammation,Perforated tympanic	CT scan
user5	Hearing loss,Facial palsy,Intense ear pain	Herpes zoster oticus,Vestibular schwannoma	Adhesive otitis media,Exostoses and osteomas	Negative Rinne test in the affected ear,Weber test lateralizes to the affected	Fluid behind the tympanic membrane,Perforated tympanic membrane	MRI
user1	Hearing loss,Intense ear pain,Ear	Herpes zoster oticus,Vestibular	Adhesive otitis media,Exostoses and	No lateralization on Weber test,Rinne	No signs of tympanic membrane perforation	CT scan
user3	Hearing loss,Swelling behind the ear,Intense ear pain,Ear discharge	Vestibular schwannoma	Acute otitis media,Adhesive otitis media,Exostoses and osteomas	Positive Rinne test in the affected ear,Weber test lateralizes to the affected	Fluid behind the tympanic membrane	No additional investigations needed
user2	Hearing loss,Intense ear pain	Vestibular schwannoma,Presbyacusis	I don't know	Negative Rinne test in the affected ear,Weber test lateralizes to the affected	No signs of tympanic membrane inflammation,Perforated tympanic	I don't know

Post-test results

User	Which of the following are potential differential diagnoses for cholesteatoma? (select all that apply)	Patient shows ear discharge, hearing loss, and intense ear pain. What could be the diagnosis? (select all that apply)	Which symptoms are associated with vestibular schwannoma (select all that apply)	What investigations are NOT required for diagnosing cholesteatoma? (select all that apply)	What would otoscopy reveal in a patient with vestibular schwannoma? (select all that apply)	What do Weber and Rinne tests results NOT show in acute otitis media? (select all that apply)
user6	Lateral skull base fractures,Squamous cell carcinoma	Acute otitis media,Inflammation of the auricle	Hearing loss,Facial numbness,Ear pain,Ear discharge,Balance issues	X-ray,Ultrasound	Normal tympanic membrane	Positive Rinne test in the affected ear,No lateralization on Weber test
user5	Acute otitis media,Herpes zoster oticus,Squamous cell carcinoma	Acute otitis media	Hearing loss,Balance issues	CT scan,X-ray,Ultrasound	Normal tympanic membrane,Granulation tissue is presented in the ear canal	Positive Rinne test in the affected ear,No lateralization on Weber test
user1	Acute otitis media,Chronic Otitis Media with Effusion	Acute otitis media	Facial numbness,Balance issues	X-ray,Ultrasound	Granulation tissue is presented in the ear canal	No lateralization on Weber test,Rinne test positive on both sides
user3	Acute otitis media,Chronic Otitis Media with Effusion,Lateral skull base fractures	Dental abscess,Inflammation of the auricle	Hearing loss,Facial numbness,Balance issues	CT scan,X-ray	Granulation tissue is presented in the ear canal	Negative Rinne test in the affected ear,No lateralization on Weber test
user2	Acute otitis media,Herpes zoster oticus,Squamous cell carcinoma	Acute otitis media	Hearing loss,Facial numbness	X-ray,Ultrasound	Granulation tissue is presented in the ear canal	Positive Rinne test in the affected ear,No lateralization on Weber test

Figure C.1: Responses to the pre-test (upper table) and the post-test (lower table).

D. Appendix

User experience questionnaire results

This application was useful overall.	I learnt something while using this application.	This application helped me with some knowledge gaps.	This application was engaging.	I like the interface of this application.	The interface of this application was clear and easy to understand.	The feedback given by this application was useful.	The feedback given by this application was informative.
Agree	Agree	Agree	Neither agree nor disagree	Agree	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree
Agree	Agree	Agree	Strongly agree	Neither agree nor disagree	Agree	Agree	Agree
Agree	Agree	Agree	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree	Neither agree nor disagree
Agree	Agree	Agree	Agree	Neither agree nor disagree	Agree	Agree	Agree
Agree	Neither agree nor disagree	Neither agree nor disagree	Agree	Neither agree nor disagree	Agree	Agree	Agree

Figure D.1: Responses to the first set of eight questions from the user experience questionnaire.

The feedback given by this application helped me progress through the exercises.	I feel that the exercises helped me learn Ear diagnoses.	I liked how the exercises were presented.	The feedback helped me understand the relationships between otology entities.	How would you rate your overall experience with the application? - Group	How would you rate your overall experience with the application?	What do you like most about the application?	What would you change or improve about the application?
Neither agree nor dis	Agree	Agree	Neither agree nor disagree	Detractor	6	That it gives you feedback	More feedback and there are some loopes in the application
Agree	Strongly agree	Neither agree nor disagree	Agree	Passive	8	Helping tool for making a plan	Some bugs: when you refresh the page you get a new story
Agree	Neither agree nor disagree	Agree	Agree	Detractor	6	The fact that the cases were generated so that the student can learn continuously new things.	The options came all if you backspaced a couple of times. This decreased the learning effect. It would be more effective if the app could read what we all typed and afterwards asked if we mean by example 'otoscopy' in that way we will be forced to think first and afterwards corrected.
Agree	Agree	Agree	Agree	Passive	7	The variety of cases, the feedback and hints were very helpful	More interactivity, pictures, more modern interface
Agree	Agree	Agree	Agree	Passive	7	Feedback buttons and how the exercise was presented	Fix a bug, when erase examination name and press backspace once more, the options for examinations appear

Figure D.2: Responses to the second set of eight questions from the user experience questionnaire, including the NPS and open-ended feedback.

E. Appendix

Utrecht University Ethics and Privacy Quick Scan

Response Summary:

Section 1. Research projects involving human participants

P1. Does your project involve human participants? This includes for example use of observation, (online) surveys, interviews, tests, focus groups, and workshops where human participants provide information or data to inform the research. If you are only using existing data sets or publicly available data (e.g. from Twitter, Reddit) without directly recruiting participants, please answer no.

- Yes

Recruitment

P2. Does your project involve participants younger than 18 years of age?

- No

P3. Does your project involve participants with learning or communication difficulties of a severity that may impact their ability to provide informed consent?

- No

P4. Is your project likely to involve participants engaging in illegal activities?

- No

P5. Does your project involve patients?

- No

P6. Does your project involve participants belonging to a vulnerable group, other than those listed above?

- No

P8. Does your project involve participants with whom you have, or are likely to have, a working or professional relationship: for instance, staff or students of the university, professional colleagues, or clients?

- No

Informed consent

PC1. Do you have set procedures that you will use for obtaining informed consent from all participants, including (where appropriate) parental consent for children or consent from legally authorized representatives? (See suggestions for information sheets and consent forms on the website.)

- Yes

PC2. Will you tell participants that their participation is voluntary?

- Yes

PC3. Will you obtain explicit consent for participation?

- Yes

PC4. Will you obtain explicit consent for any sensor readings, eye tracking, photos, audio, and/or video recordings?

- Not applicable

PC5. Will you tell participants that they may withdraw from the research at any time and for any reason?

- Yes

PC6. Will you give potential participants time to consider participation?

- Yes

PC7. Will you provide participants with an opportunity to ask questions about the research before consenting to take part (e.g. by providing your contact details)?

- Yes

PC8. Does your project involve concealment or deliberate misleading of participants?

- No

Section 2. Data protection, handling, and storage

The General Data Protection Regulation imposes several obligations for the use of **personal data** (defined as any information relating to an identified or identifiable living person) or including the use of personal data in research.

D1. Are you gathering or using personal data (defined as any information relating to an identified or identifiable living person)?

- No

Section 3. Research that may cause harm

Research may cause harm to participants, researchers, the university, or society. This includes when technology has dual-use, and you investigate an innocent use, but your results could be used by others in a harmful way. If you are unsure regarding possible harm to the university or society, please discuss your concerns with the Research Support Office.

H1. Does your project give rise to a realistic risk to the national security of any country?

- No

H2. Does your project give rise to a realistic risk of aiding human rights abuses in any country?

- No

H3. Does your project (and its data) give rise to a realistic risk of damaging the University's reputation? (E.g., bad press coverage, public protest.)

- No

H4. Does your project (and in particular its data) give rise to an increased risk of attack (cyber- or otherwise) against the University? (E.g., from pressure groups.)

- No

H5. Is the data likely to contain material that is indecent, offensive, defamatory, threatening, discriminatory, or extremist?

- No

H6. Does your project give rise to a realistic risk of harm to the researchers?

- No

H7. Is there a realistic risk of any participant experiencing physical or psychological harm or discomfort?

- No

H8. Is there a realistic risk of any participant experiencing a detriment to their interests as a result of participation?

- No

H9. Is there a realistic risk of other types of negative externalities?

- No

Section 4. Conflicts of interest

C1. Is there any potential conflict of interest (e.g. between research funder and researchers or participants and researchers) that may potentially affect the research outcome or the dissemination of research findings?

- No

C2. Is there a direct hierarchical relationship between researchers and participants?

- No

Section 5. Your information.

This last section collects data about you and your project so that we can register that you completed the Ethics and Privacy Quick Scan, sent you (and your supervisor/course coordinator) a summary of what you filled out, and follow up where a fuller ethics review and/or privacy assessment is needed. For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information. Please see the guidance on the ICS Ethics and Privacy website on what happens on submission.

Z0. Which is your main department?

- Information and Computing Science

Z1. Your full name:

Andrei Marinin

Z2. Your email address:

a.a.marinin@students.uu.nl

Z3. In what context will you conduct this research?

- As a student for my master thesis, supervised by::
  s.a.sosnovsky@uu.nl

Z5. Master programme for which you are doing the thesis

- Human-Computer Interaction

Z6. Email of the course coordinator or supervisor (so that we can inform them that you filled this out and provide them with a summary):

s.a.sosnovsky@uu.nl

Z7. Email of the moderator (as provided by the coordinator of your thesis project):

m.m.a.degraaf@uu.nl

Z8. Title of the research project/study for which you filled out this Quick Scan:

From Practice to Proficiency: Evaluating the Educational Value of a Diagnostic Process Simulation App for Medical Students

Z9. Summary of what you intend to investigate and how you will investigate this (200 words max):

This thesis investigates ontology-driven approaches to develop an innovative medical training application, addressing the demand for interactive, immersive educational tools in the medical field. The main challenge is the scarcity of engaging training tools that mirror the complexities of real-world clinical scenarios.
The thesis proposes combining Human-Computer Interaction (HCI) capabilities with the structured content of medical ontologies to create an application that generates medical exercises and case studies in a user-friendly interface and provides useful feedback on user's actions. This aims to offer a more immersive learning experience. The proposed application addresses this by dynamically generating varied exercises, enhancing learning efficiency and effectiveness. The research methodology includes developing a prototype application and testing it with medical students to assess its effectiveness, engagement, and educational value. Assessments involve surveys, interviews, and performance tracking. The anticipated outcome is a notable improvement in medical education quality through enhanced interactive learning experiences. This thesis is expected to provide empirical evidence supporting the integration of HCI and ontologies in educational tools, setting a new precedent in medical education technology.

Z10. In case you encountered warnings in the survey, does supervisor already have ethical approval for a research line that fully covers your project?

- Not applicable

Scoring

- Privacy: 0
- Ethics: 0

Bibliography

[1] P. Khanna, C. Roberts, and A. S. Lane, "Designing health professional education curricula using systems thinking perspectives," *BMC Medical Education*, vol. 21, pp. 1–8, 2021.

[2] L. M. Buja, "Medical education today: All that glitters is not gold," *BMC medical education*, vol. 19, pp. 1–11, 2019.

[3] G. Alinier and D. Oriot, "Simulation-based education: Deceiving learners with good intent," *Advances in Simulation*, vol. 7, no. 1, p. 8, 2022.

[4] G. Kurdi, J. Leo, B. Parsia, U. Sattler, and S. Al-Emari, "A systematic review of automatic question generation for educational purposes," *International Journal of Artificial Intelligence in Education*, vol. 30, pp. 121–204, 2020.

[5] S. Abd El-Rahman and A. H. Zolait, "Automated test paper generation using utility based agent and shuffling algorithm," *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, vol. 14, no. 1, pp. 69–83, 2019.

[6] O. Westphal, "A framework for generating diverse haskell-i/o exercise tasks," in *International Workshop on Functional and Constraint Logic Programming*, Springer, 2020, pp. 97–114.

[7] X. Yao, G. Bouma, and Y. Zhang, "Semantics-based question generation and implementation," *Dialogue and Discourse*, vol. 3, no. 2, pp. 11–42, 2012.

[8] B. Das, M. Majumder, S. Phadikar, and A. A. Sekh, "Automatic question generation and answer assessment: A survey," *Research and Practice in Technology Enhanced Learning*, vol. 16, no. 1, pp. 1–15, 2021.

[9] M. Ivanović and Z. Budimac, "An overview of ontologies and data resources in medical domains," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5158–5166, 2014.

[10] R. Hoehndorf, P. N. Schofield, and G. V. Gkoutos, "Analysis of the human diseasome using phenotype similarity between common, genetic and infectious diseases," *Scientific Reports*, vol. 5, no. 1, pp. 1–14, 2015.

[11] S. Harispe, D. Sánchez, S. Ranwez, S. Janaqi, and J. Montmain, "A framework for unifying ontology-based semantic similarity measures: A study in the biomedical domain," *Journal of biomedical informatics*, vol. 48, pp. 38–53, 2014.

[12] O. Bodenreider and R. Stevens, "Bio-ontologies: Current trends and future directions," *Briefings in Bioinformatics*, vol. 7, no. 3, pp. 256–274, 2006.

[13] B. Liu, K. Huang, J. Li, and M. Zhou, "An incremental and distributed inference method for large-scale ontologies based on mapreduce paradigm," *IEEE transactions on cybernetics*, vol. 45, no. 1, pp. 53–64, 2014.

[14] A. Malhotra, E. Younesi, M. G¨undel, B. M¨uller, M. T. Heneka, and M. Hofmann-Apitius, "Ado: A disease ontology representing the domain

knowledge specific to alzheimer's disease," *Alzheimer's & dementia*, vol. 10, no. 2, pp. 238–246, 2014.

[15] B. Smith, S. Arabandi, M. Brochhausen, *et al.*, "Biomedical imaging ontologies: A survey and proposal for future work," *Journal of pathology informatics*, vol. 6, no. 1, p. 37, 2015.

[16] S. Zhang and O. Bodenreider, "Experience in aligning anatomical ontologies," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 3, no. 2, pp. 1–26, 2007.

[17] J. B. Bard and S. Y. Rhee, "Ontologies in biology: Design, applications and future challenges," *Nature Reviews Genetics*, vol. 5, no. 3, pp. 213–222, 2004.

[18] D. G. Sampson, M. D. Lytras, G. Wagner, and P. Diaz, "Ontologies and the semantic web for e-learning (guest editorial)," *Journal of Educational Technology & Society*, vol. 7, no. 4, pp. 26–28, 2004.

[19] J. Wang, T. Mendori, and J. Xiong, "A language learning support system using course-centered ontology and its evaluation," *Computers & Education*, vol. 78, pp. 278–293, 2014.

[20] B. Chen, C.-Y. Lee, and I.-C. Tsai, "Ontology-based e-learning system for personalized learning," *International Journal of Innovation, Management and Technology*, vol. 3, no. 4, p. 464, 2012.

[21] J. Hattie and H. Timperley, "The power of feedback," *Review of educational research*, vol. 77, no. 1, pp. 81–112, 2007.

[22] V. J. Shute, "Focus on formative feedback," *Review of educational research*, vol. 78, no. 1, pp. 153–189, 2008.

[23] R. B. Clariana, "Differential memory effects for immediate and delayed feedback: A delta rule explanation of feedback timing effects.," 1999.

[24] M. J. Gierl and H. Lai, "Using automatic item generation to create solutions and rationales for computerized formative testing," *Applied psychological measurement*, vol. 42, no. 1, pp. 42–57, 2018.

[25] C. Gutl, K. Lankmayr, J. Weinhofer, and M. Hofler, "Enhanced automatic question creator–eaqc: Concept, development and evaluation of an automatic test item creation tool to foster modern e-education.," *Electronic Journal of e-Learning*, vol. 9, no. 1, pp. 23–38, 2011.

[26] M. N. Demaidi, M. M. Gaber, and N. Filer, "Ontopefege: Ontology-based personalized feedback generator," *IEEE Access*, vol. 6, pp. 31 644–31 664, 2018.

[27] P. E. Antoniou, E. Chondrokostas, C. Bratsas, P.-M. Filippidis, and P. D. Bamidis, "A medical ontology informed user experience taxonomy to support co-creative workflows for authoring mixed reality medical education spaces," in *2021 7th International Conference of the Immersive Learning Research Network (iLRN)*, IEEE, 2021, pp. 1–9.

[28] J. Leo, G. Kurdi, N. Matentzoglu, *et al.*, "Ontology-based generation of medical, multi-term mcqs," *International Journal of Artificial Intelligence in Education*, vol. 29, pp. 145–188, 2019.

[29] M. Radovic, N. Petrovic, and M. Tosic, "An ontology-driven learning assessment using the script concordance test," *Applied Sciences*, vol. 12, no. 3, p. 1472, 2022.

[30] C. Bratsas, E. Kaimakamis, V. Koutkias, P. D. Bamidis, and N. Maglaveras, "An ontology-based approach to constructing medical computational problems for use in electronic medical education," in *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*, IEEE, 2007, pp. 669–674.

[31] H. Y. So, P. P. Chen, G. K. C. Wong, and T. T. N. Chan, "Simulation in medical education," *Journal of the Royal College of Physicians of Edinburgh*, vol. 49, no. 1, pp. 52–57, 2019.

[32] D. R. Danforth, M. Procter, R. Chen, M. Johnson, and R. Heller, "Development of virtual patient simulations for medical education," *Journal For Virtual Worlds Research*, vol. 2, no. 2, 2009.

[33] A. H. Al-Elq, "Simulation-based medical teaching and learning," *Journal of family and Community Medicine*, vol. 17, no. 1, pp. 35–40, 2010.

[34] A. Gayef, "Using simulated patients in medical and health professions education," in *SHS Web of Conferences*, EDP Sciences, vol. 66, 2019, p. 01 016.

[35] D. Pai, "Use of simulation for undergraduate medical education," *International Journal of Advanced Medical and Health Research*, vol. 5, no. 1, pp. 3–6, 2018.

[36] L. Rundo, R. Pirrone, S. Vitabile, E. Sala, and O. Gambino, "Recent advances of hci in decision-making tasks for optimized clinical workflows and precision medicine," *Journal of biomedical informatics*, vol. 108, p. 103 479, 2020.

[37] I. Healthcare. "Isabel ddx companion." Accessed: 2024-07-09. (2024), [Online]. Available: `https://www.isabelhealthcare.com/products/isabel-ddx-companion`.

[38] VisualDx. "Visualdx - clinical decision support." Accessed: 2024-07-09. (2024), [Online]. Available: `https://www.visualdx.com`.

[39] M. G. Hospital. "Dxplain - decision support system." Accessed: 2024-07-09. (2024), [Online]. Available: `http://dxplain.org`.

[40] InSimu. "Insimu patient - virtual patient simulator." Accessed: 2024-07-09. (2024), [Online]. Available: `https://www.insimu.com`.

[41] M. Joyworks. "Prognosis: Your diagnosis - clinical case simulations." Accessed: 2024-07-09. (2024), [Online]. Available: `https://www.medicaljoyworks.com/prognosis-your-diagnosis`.

[42] C. by Clinical Odyssey. "Case - daily clinical cases for medical professionals." Accessed: 2024-07-09. (2024), [Online]. Available: `https://www.clinicalodyssey.com`.

[43] S. C. for Biomedical Informatics Research. "Protégé - free, open-source ontology editor and framework." Accessed: 2024-07-09. (2024), [Online]. Available: `https://protege.stanford.edu`.