UNIVERISTY OF PIRAEUS - DEPARTMENT OF INFORMATICS

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ - ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

MSc «AI-Based Model for Knowledge Specific Assistance»

ΠΜΣ «Μοντέλο Τεχνητής Νοημοσύνης για Βοήθεια σε Συγκεκριμένη Γνώση»

MSc Thesis

Μεταπτυχιακή Διατριβή

Thesis Title: Τίτλος Διατριβής:	**AI-Based Model for Knowledge Specific Assistance** Μοντέλο Τεχνητής Νοημοσύνης για Βοήθεια σε Συγκεκριμένη Γνώση
Student's name-surname: Ονοματεπώνυμο φοιτητή:	**Thanos Apostolou** Θάνος Αποστόλου
Father's name: Πατρώνυμο:	**Christos** Χρήστος
Student's ID No: Αριθμός Μητρώου:	**MPSP2203** ΜΠΣΠ2203
Supervisor: Επιβλέπων:	**Dionisios Sotiropoulos, Assistant Professor** Διονύσιος Σωτηρόπουλος, Επίκουρος Καθηγητής

September 2024/ Σεπτέμβριος 2024

3-Member Examination Committee		
Τριμελής Εξεταστική Επιτροπή		
Dionisios Sotiropoulos **Assistant Professor** Διονύσιος Σωτηρόπουλος Επίκουρος Καθηγητής		

Contents

Abstract

This MSc thesis is about utilizing artificial intelligence models in order to find specific knowledge. As part of this goal we will develop a complete web application, where users will be able to ask questions to artificial intelligence models, which will answer them based on a specific context. We will follow two different methodologies. For the first methodology we will create our own text generation AI model [1] which will be trained to understand specific knowledge. For the second methodology, we will use existing artificial intelligence models, trying to limit them so that they respond only to the specific knowledge context that we have chosen. In the end we will be able to come to conclusions about the usefulness of these methodologies.

Περίληψη

Η παρούσα μεταπτυχιακή εργασία ασχολείται με την αξιοποίηση μοντέλων τεχνητής νοημοσύνης για την υποβοήθηση ανεύρεσης συγκεκριμένης γνώσης. Στα πλαίσια αυτού του στόχου θα αναπτύξουμε μια πλήρη διαδικτυακή εφαρμογή, στην οποία οι χρήστες θα μπορούν να κάνουν ερωτήσεις σε μοντέλα τεχνητής νοημοσύνης, τα οποία θα τους απαντάνε με βάση συγκεκριμένο πλαίσιο. Θα ακολουθήσουμε δύο διαφορετικές μεθοδολογίες. Για την πρώτη μεθοδολογία θα δημιουργήσουμε ένα δικό μας μοντέλο τεχνητής νοημοσύνης παραγωγής κειμένου [1] το οποίο θα εκπαιδευτεί για να κατανοεί συγκεκριμένη γνώση. Για την δεύτερη μεθοδολογία θα χρησιμοποιήσουμε υπάρχοντα μοντέλα τεχνητής νοημοσύνης προσπαθώντας να τα περιορίσουμε ώστε να απαντάνε μόνο στο συγκεκριμένο πλαίσιο γνώσης που έχουμε επιλέξει. Στο τέλος θα μπορέσουμε να καταλήξουμε σε συμπεράσματα. Στο τέλος θα μπορέσουμε να καταλήξουμε σε συμπεράσματα για την χρησιμότητα αυτών των μεθοδολογιών. [2]

1. Introduction

In our era, the knowledge we have acquired is bigger than ever. The number of books, notes, web pages and other forms of content keeps increasing year by year. It is impossible for any human being, to be able to read an process all this available knowledge. Fortunately, technology has been greatly improved and is being used daily for tasks involving knowledge search and analysis. While traditional tools like search engines made it easier for us to find existing knowledge, in the past years we have observed the increasing development of tools using artificial intelligence. We will study the usage of text generation machine learning models in specific knowledge search and analysis assistance. We will use two different methodologies for these tasks and we will develop a full web application with which users will be able to ask questions

In chapter 2 we will describe and analyze the fundamental theoretical concepts needed for better understanding of this thesis. We will also describe the various technologies and their advantage, which we will use for our application development and deployment.

In chapter 3 we will dive in the details of the two methodologies that we will use. We will compare them and we will describe their advantages and disadvantages.

In chapter 4 we will describe the architecture and the implementation of our application. We will show the components which construct our application, the tasks each component can perform and how they are connected together.

In chapter 5 we will show the design and execution results of our deployed application. We will investigate the various ways in which our application can be used by the users in order to find specific knowledge based on raw data like documents or web pages.

In chapter 6 we will write our conclusions we reached. We will describe the problems and limitations we faced. Finally, we will specify future improvements that can be made as well as future goals about scaling and expand the core idea.

2. Theory and Literature Review

In this chapter we will talk about the theoretic terms that this thesis is based upon. We will also describe the main technologies which we will use.

2.1. Theoretic Terms

2.1.1. Artificial intelligence

In the general sense, Artificial intelligence (AI) is intelligence exhibited by machines, particularly computer systems. It is a field of research in computer science that develops and studies methods and software that enable machines to perceive their environment and use learning and intelligence to take actions that maximize their chances of achieving defined goals. Such machines may be called AIs. [3]

Intelligence can be considered to be a property of internal thought processes and reasoning, or a property of intelligent behavior, an external characterization. From these two dimensions (human vs. rational and thought vs. behavior) there are four possible combinations. The methods used are necessarily different: the pursuit of human-like intelligence must be in part an empirical science related to psychology, involving observations and hypotheses about actual human behavior and thought processes; a rationalist approach, on the other hand, involves a combination of mathematics and engineering, and connects to statistics, control theory, and economics. These 4 approaches are the following:[2]

- Acting humanly: The Turing test approach

    The Turing test, proposed by Alan Turing (1950) and it consists of 4 core principles that a computer would need to follow in order to pass it.
    ‣ natural language processing to communicate successfully in a human language
    ‣ knowledge representation to store what it knows or hears
    ‣ automated reasoning to answer questions and to draw new conclusions
    ‣ machine learning to adapt to new circumstances and to detect and extrapolate patterns

    The full turing test is completed with 2 additional characteristics which have been added by later researchers:
    ‣ computer vision and speech recognition to perceive the world
    ‣ robotics to manipulate objects and move about

- Thinking humanly: The cognitive modeling approach

    We can determine if a computer or a program thinks like a human by analyzing the human thought in 3 main concepts:
    ‣ introspection - trying to catch our own thoughts as they go by
    ‣ psychological experiments - observing a person in action
    ‣ brain imaging - observing the brain in action

- Thinking rationally: The "laws of thought" approach

    Rationally thinking can be achieved by following the rules defined by the "logic" study field. When conventional logic requires knowledge that cannot be obtained realistically, then the theory of probability helps us define logical thinking.

- Acting rationally: The rational agent approach

Rational thinking can achieve a construction of a comprehensive model of rational thought, but cannot generate intelligent behavior by itself. A rational agent is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome.

2.1.2. Machine Learning Models

We described the fundamental concepts with which artificial intelligence is defined.
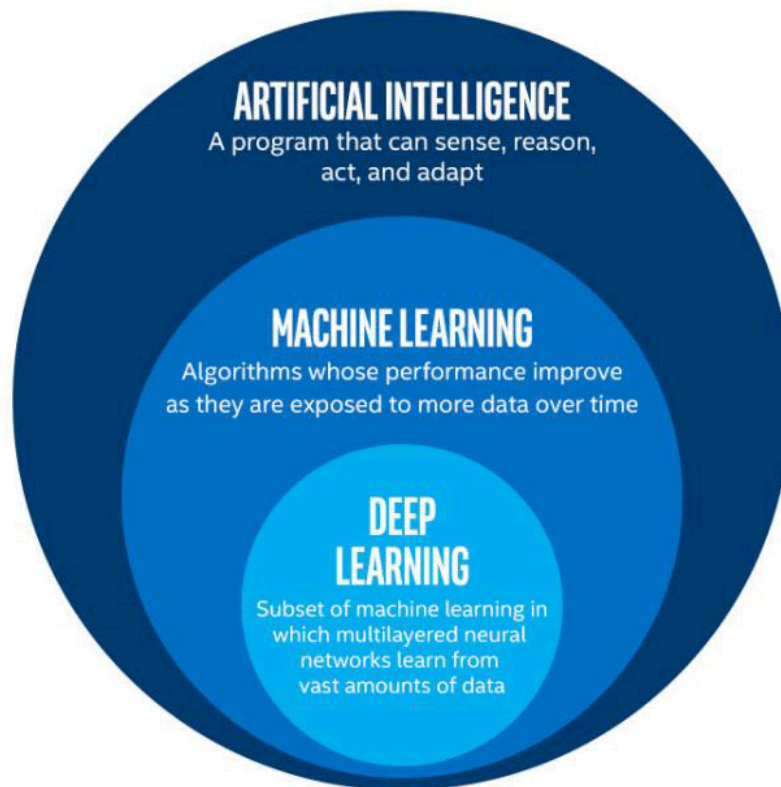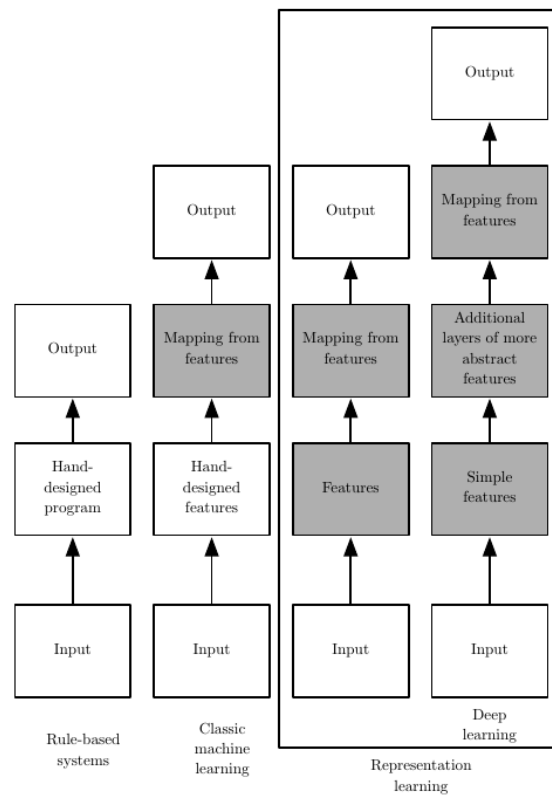

IMAGE 1: Venn Diagram for AI, ML, Deep Learning [4]

IMAGE 2: Flowcharts showing how the different parts of an AI system relate to each other within different AI disciplines. Shaded boxes indicate components that are able to learn from data. [5]

2.2. Technologies

2.2.1. Programming Languages

2.2.2. Libraries

2.2.3. Containers, Docker and Kubernetes

3. Specific Knowledge Assistance Approaches

3.1. Custom Text Generation Model Method

3.2. Retrieval Augmented Generation Method

4. System Architecture

5. Usage and Execution of the Application

6. Conclusions and Future Work

Bibliography

[1] H. Face, "Text Generation." [Online]. Available: https://huggingfagce.co/tasks/text-generation

[2] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th Global Edition. Pearson, 2021.

[3] Wikipedia, "Artificial intelligence." [Online]. Available: https://en.wikipedia.org/wiki/Artificial_intelligence

[4] R. Khalkar, A. S. Dikhit, and A. Goel, "Handwritten Text Recognition using Deep Learning (CNN & RNN)," *International Advanced Research Journal in Science, Engineering and Technology*, 2021, [Online]. Available: https://www.researchgate.net/publication/353939315_Handwritten_Text_Recognition_using_Deep_Learning_CNN_RNN

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Artificial Intelligence: Deep learning - adaptive computation and machine learning*. 2016.