



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

**SCHOOL OF SCIENCE
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS**

**INTERDISCIPLINARY MSc PROGRAM
LANGUAGE TECHNOLOGY**

MSc THESIS

**Human and ChatGPT perspectives on literature:
A comparative analysis of Greek high school literary
texts**

Kalliopi S. Kakamouka

Supervisor: Vasilios Papavasiliou, Scientific Assistant at ILSP "Athena"

ATHENS

December 2024

MSc THESIS

Human and ChatGPT perspectives on literature:
A comparative analysis of Greek high school literary texts

Kalliopi S. Kakamouka

S.N.: 7115182200030

SUPERVISOR: **Vasilios Papavasiliou**, Scientific Assistant at ILSP “Athena”

EXAMINATION **Sokratis Sofianopoulos**, Scientific Assistant at ILSP “Athena”

COMMITTEE: **Spyridoula Stamouli**, Specialised Scientific Personnel at ILSP RC “Athena”

ATHENS

December 2024

ABSTRACT

This master's thesis investigates the comparative analysis of human and ChatGPT responses to 232 literary texts, which are included in the curriculum of the final two years of Greek high school, as established by the Greek Ministry of Education. Each text was paired with its corresponding literary questions and human-generated answers, sourced from the Ministry or supplemented by the author. ChatGPT was tasked with answering the same questions, and its responses were compared to human answers based on predefined literary key points, including textual markers, figurative language, and subjective interpretation.

To support future research, the bundles were converted into XML files, creating a structured dataset for potential use in training or fine-tuning language models (the XML files and the rest of the dataset can be found in this address: https://drive.google.com/drive/folders/1L6yNAI_kL3kVnISTGcRqkyFO7zvH4fn_r?usp=sharing). Additionally, a brief experiment with Meltemi, the first open-source Greek LLM, was conducted to assess its capabilities on a subset of the dataset.

The results highlight ChatGPT's strengths in adherence to word count and subjective expression but reveal its limitations in capturing nuanced literary features. In contrast, human responses excel in contextual depth and richness, locating suitable literary features, but comparing to the ChatGPT language model seem to struggle to stay within the requested word count.

Meltemi shows promise but struggles staying within the required word count, while it seems to be able to locate very interesting literary elements within each text and thus provide complete answers.

This study not only evaluates LLMs' potential but also provides a comparative framework for analyzing AI and human performance in literary analysis of Greek-language educational contexts.

SUBJECT AREA: Natural Language Processing

KEYWORDS: LLM, NLP, ChatGPT, Meltemi, XML, literary analysis

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία ερευνά την συγκριτική ανάλυση ανθρώπινων και παραγόμενων από το ChatGPT απαντήσεων σε 232 λογοτεχνικά κείμενα, τα οποία αποτελούν ύλη των τελευταίων δύο τάξεων του λυκείου και παρέχονται από το Υπουργείο Παιδείας. Κάθε κείμενο συνοδεύεται από τις αντίστοιχες λογοτεχνικές του ερωτήσεις καθώς και τις απαντήσεις που έχουν δοθεί από άνθρωπο, οι οποίες προέρχονται είτε από το ίδιο το Υπουργείο Παιδείας ή συμπληρώνονται και από την ίδια την συγγραφέα. Στο ChatGPT τέθηκαν οι ίδιες ερωτήσεις και οι απαντήσεις του συγκρίθηκαν με αυτές που δόθηκαν από άνθρωπο βάσει προκαθορισμένων σημείων-κλειδιών, συμπεριλαμβανομένων στοιχείων/παραδειγμάτων μέσα από το κείμενο, των κειμενικών δεικτών, των εκφραστικών μέσων καθώς και της προσωπικής άποψης.

Για την υποστήριξη μελλοντικών ερευνών, οι απαντήσεις που δόθηκαν από τον άνθρωπο μετατράπηκαν σε αρχεία XML, δημιουργώντας ένα δομημένο σύνολο δεδομένων για πιθανή χρήση, με στόχο την εκπαίδευση γλωσσικών μοντέλων (τα αρχεία XML καθώς και το υπόλοιπο σύνολο δεδομένων βρίσκεται στον ακόλουθο σύνδεσμο: https://drive.google.com/drive/folders/1L6yNAI_kL3kVnIStGcRqkyFO7zvH4fn_r?usp=sharing). Επιπλέον, πραγματοποιήθηκε ένα σύντομο πείραμα με το Μελέμι, το πρώτο ελληνικό LLM ανοιχτού κώδικα, με στόχο την εκτίμηση των δυνατοτήτων του σε ένα υποσύνολο του συνόλου δεδομένων.

Τα αποτελέσματα υπογραμμίζουν τα δυνατά σημεία του ChatGPT που αφορούν την παραμονή των απαντήσεων του εντός προκαθορισμένων ορίων λέξεων, καθώς και την έκφραση της “υποκειμενικής” του άποψης, αλλά και τους περιορισμούς του στην εύρεση λογοτεχνικών στοιχείων. Από την άλλη μεριά, οι απαντήσεις που δόθηκαν από τον άνθρωπο τα πηγαίνουν εξαιρετικά όσον αφορά την κατανόηση του βάθους και της πολυπλοκότητας του περιεχομένου, εντοπίζοντας τα κατάλληλα λογοτεχνικά στοιχεία, ενώ σε αντίθεση με το γλωσσικό μοντέλο ChatGPT, αδυνατούν να επιτύχουν τα απαιτούμενα όρια λέξεων.

Το Μελέμι φαίνεται να έχει πολλές δυνατότητες, αλλά κι αυτό με την σειρά του δυσκολεύεται να μείνει εντός ορίων, ενώ φαίνεται να εντοπίζει

ενδιαφέροντα λογοτεχνικά στοιχεία στο κείμενο, παρέχοντας ολοκληρωμένες απαντήσεις.

Αυτή η έρευνα δεν αξιολογεί μονάχα τις δυνατότητες των γλωσσικών μοντέλων, αλλά παρέχει και ένα συγκριτικό πλαίσιο αναλύοντας την τεχνητή νοημοσύνη και την ανθρώπινη απόδοση στην λογοτεχνική ανάλυση εκπαιδευτικών συγκειμένων ελληνικής γλώσσας.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Επεξεργασία Φυσικής Γλώσσας

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: LLM, NLP, ChatGPT, Meltemi, XML, λογοτεχνική ανάλυση

To my family, who always supports my dreams

Contents

List of Figures, Tables and Graphs	9
List of Figures	9
List of Tables	9
List of Graphs	9
Preface	10
0. Introduction	11
0.1. the goal of the research	11
0.2. the structure of the thesis	12
0.3. methodology	12
1. Theoretical Section	13
1.1.1. the importance of language as a communication tool	13
1.1.2. language and technology	14
1.1.3. the evolution of NLP	14
1.2. LLMs and ChatGPT	15
1.2.1. what they are	15
1.2.2. ChatGPT's function	16
1.3. ChatGPT in education	16
1.3.1. education and technology	16
1.3.2. the use of AI and ChatGPT by students	17
1.4. ChatGPT in different subjects	17
1.4.1. ChatGPT in mathematics and physics	17
1.4.2. ChatGPT and literature in education	20
2. Empirical Section	21
2.1. data collection and classification	22

2.2.	bundle contents with examples	23
2.2.1.	source file content	23
2.2.2.	text file question content	24
2.2.3.	human answer content	25
2.2.4.	ChatGPT’s answer content	26
2.2.5.	excel file content	27
2.3.	XML schema	31
3.	Meltemi	36
3.1.	introduction to Meltemi	36
3.2.	experimenting with Meltemi	36
4.	Results	37
4.1.1.	occurrence count of required key features (human, ChatGPT) 37	
4.1.2.	occurrence count of required key features (Meltemi)	38
4.2.	occurrence count of “ΔZ” in required key features	40
4.3.	ChatGPT’s most referenced textual markers/figurative language 41	
5.	Conclusion	43
	Terminology Table	45
	Abbreviations – Acronyms – Initialisms	46
	References	47

List of Figures, Tables and Graphs

List of Figures

Figure 1: Excel file structure	29
Figure 2: Custom XML schema	31

List of Tables

Table 1: ChatGPT's mathematical capabilities according to Rane, N., [25]....	18
Table 2: Bundle IDs and Ministry IDs.....	23
Table 3: Expected features and corresponding examples	28
Table 4: Interpretation of the structure and the content of an XML file	31

List of Graphs

Graph 1: Comparison of Key Feature Occurrences in Human and ChatGPT Responses.....	37
Graph 2: Key Feature Occurrences in Meltemi's Responses	39
Graph 3: Frequency of Non-Required Features in Answers (Marked with "ΔZ")	40
Graph 4: ChatGPT's most referenced textual markers & figurative language	41
Graph 5: Meltemi's most referenced textual markers & figurative language...	42

Preface

This master's thesis was completed in Athens, Greece during the academic year of 2023-2024 as part of the interdisciplinary Master program "Language Technology" at the faculty of Informatics and Telecommunications, at the National and Kapodistrian University of Athens.

Supervised by Mr. Vasilios Papavasiliou, Scientific Associate at the Institute for language and Speech Processing "Athena" and co-advised by Mr. Sokratis Sofianopoulos and Ms. Spyridoula Stamouli, also Scientific Associates at the Institute for Language and Speech Processing "Athena", this thesis examined, analyzed and collected data from responses provided by both humans and the ChatGPT 3.5 language model, based on various literary texts from the last two years off the Greek high school curriculum.

At this point I would like to extend my sincere gratitude to my supervisor, Vasilios Papavasiliou, for his unwavering support and invaluable guidance throughout the conducting of my master's thesis, and whose expertise and encouragement were instrumental in shaping this work.

0. Introduction

The intersection of artificial intelligence and education has opened new possibilities for enhancing learning and evaluation processes. According to Dash and Bhoi [1], artificial intelligence (AI) is rapidly changing the way we learn and teach. From personalized tutoring to intelligent content creation, AI is transforming education.

Among the numerous developments in artificial intelligence, Large Language Models have surfaced as influential instruments with the ability to generate detailed responses across a variety of domains. One emerging AI tool with potential applications in education is ChatGPT. The model's ability to generate answers in various subjects is quite remarkable. As mentioned by Azaria et al. [2] "ChatGPT serves as a valuable tool not only in the field of research, but also in education, as it can provide responses to numerous questions, while creating dynamic, flexible and customized content."

0.1. the goal of the research

This thesis focuses on collecting data from the domain of literature, such as literary texts and their corresponding questions, but also required literary key features, which are provided by the Greek ministry of Education and are taught in the last two years of Greek high school.

The goal is to explore ChatGPT's capabilities in literary analysis based on the collected literary texts, while comparing its generated responses to human-given answers, which serve as the ground truth. By evaluating both responses to these literary questions, the study examines the potential and limitations of the model in the subject of literature as an educational assistant.

A notable aspect of the study is the conversion of the created bundles into XML files, enabling systematic comparisons and offering a resource for future model training or fine-tuning.

Lastly, the inclusion of Meltemi, the first Greek open-source LLM, which provides insights into AI models tailored to the Greek language, has also been evaluated on several of the collected literary data, providing insightful information on its performance, in comparison to the answers provided by ChatGPT and the human factor as well.

0.2. **the structure of the thesis**

The thesis is organized into five chapters. The first chapter, titled “Theoretical Section”, presents the foundational theory underlying this research. It begins by exploring the significance and evolution of language as a tool, followed by an in-depth discussion on the role and application of Large Language Models in education.

The second chapter, "Empirical Section," focuses on the research process itself. It outlines the data collection and organization, provides a detailed examination of each file's content, and includes an analysis of experiments conducted with ChatGPT, as well as the conversion of data into XML format.

The third chapter is dedicated to Meltemi, the first Greek Large Language Model. This section presents a brief experiment designed to assess the model's capabilities by posing some of the collected literary questions.

Chapter four examines the results of the research. It includes various graphs that analyze the occurrence of key features in the responses generated by ChatGPT, Meltemi, and humans, as well as graphs highlighting the most frequently referenced textual markers and instances of figurative language identified by both models.

The final chapter offers conclusions drawn from the research. It discusses the practical performance of each model, as indicated by the generated responses, and compares them to the ground truth, providing insights into the models' effectiveness.

0.3. **methodology**

This section outlines the methodology used to conduct the research. Specifically, the study involved selecting 232 literary texts coming from a broader collection provided by the Greek Ministry of Education, which are taught in the last two years of Greek high school. The Ministry supplies answers to the literary questions formulated based on these texts, however, in cases where the answers were not provided due to their subjective nature, I, as a linguist, intervened to complete them. These answers, which are predominantly a combination of those provided by the Ministry and my own contributions, serve as the ground truth for the study.

After selecting the 232 literary texts and compiling the corresponding answers, the next step was to place the exact same questions, accompanied by their literary texts to ChatGPT 3.5.

The study followed a structured approach to data collection and organization. For each literary text, a “bundle” was created containing six components: the text, the corresponding question(s), the human response, ChatGPT’s generated response, any relevant instructions, an Excel file for comparative evaluation. These bundles ensured a consistent framework for analysis, with unique identifiers assigned to each literary text and its associated data for organizational purposes.

Additionally, a brief experiment was conducted with Meltemi, the first Greek open-source LLM. Ten of the 232 literary texts and their corresponding questions were selected to test Meltemi’s performance using the same evaluation framework. The findings from this experiment provided supplementary insights into the capabilities and limitations of Greek-specific AI models, enriching the study’s comparative analysis.

The responses were evaluated based on five key literary features: the use of textual elements and examples, identification of textual markers, application of figurative language, subjective agreement or disagreement, and adherence to word count guidelines. An Excel-based system was used to annotate and compare these features across human and ChatGPT responses, with discrepancies and patterns systematically recorded. To enable further research and facilitate the training or fine-tuning of other language models, the human responses were converted into XML files, preserving their structure and metadata.

1. Theoretical Section

1.1.1. the importance of language as a communication tool

According to Ferdinand de Saussure [\[3, p. 9\]](#), language (or *langue*) is defined as follows: "Language is a social product of the faculty of speech and a collection of necessary conventions that have been adopted by a social body to permit individuals to exercise that faculty. It is a self-contained whole and a principle of classification.". Saussure distinguishes language from speech (*langage*), stating

that language is a structured system and the social aspect of speech that individuals cannot create independently. Besides Saussure, the definition of language has been studied by many, with a few of them like Noam Chomsky [4, p. 1] stating that “language is a natural object, a component of the human mind, physically represented in the brain and part of the biological endowment of the species” and Henry Sweet [5, p. 14], who states, that “every sentence or word by which we express our ideas has a certain definite form of its own by virtue of the sounds of which it is made up, and has a more or less definite meaning”. The importance of language is indisputable, as it holds the memory and information of thousands of years of a culture. Grenoble and Whaley [6, p. 3] explain that language is an important way to maintain the association with a culture’s past and preserve its uniqueness in the present, considering that it is a visible and powerful indicator of a group’s identity.

1.1.2. **language and technology**

For thousands of years, we have used language as a communication tool, to spread and exchange ideas, cultural elements and emotions. However, with the evolution of our society, came the need for advanced technology in many different fields, affecting the way that language is communicated and exchanged. The evolution of language and technology is interdependent, and that relationship goes back all the way to the industrial revolution, where “science and technology have promoted the development of the language” [7]. This strong bond between language and technology is commented by Del Vecchio Lança, L. [8] as follows: “Language and technology constitute our world and shape our sense of reality”, whereas Bardeen and Cerpa [9, p. 2] remark that “technology and society are inexorably intertwined. The needs of society create the need for various technologies, and new technologies generate new needs for society”.

1.1.3. **the evolution of NLP**

As mentioned above, technological evolution is a correlated event to the evolution of our society and vice versa. One example of technological growth is the field of Natural Language Processing, which started to arise in the 1940’s. Considering the importance of translation during this period of time, the demand

for machines capable of translating and understanding human language became inevitable. However, NLP faced challenges highlighted by researchers like Noam Chomsky in the 1950s. From 1957 to 1970, NLP research was divided into symbolic (rule-based) and stochastic (statistical) approaches. Post-1970, new areas emerged, including logic-based paradigms and discourse modeling, which examined human-computer interactions [10]. Until the 1980s, natural language processing relied on handwritten rules for machines to understand language. The advent of machine learning brought a major shift, replacing these rules with algorithms like decision trees and probabilistic models, which could better handle ambiguity and provide evidence-based decisions. Currently, according to Johri et al. [11], deep learning dominates NLP effectively addressing the complexities and ambiguities of language that rule-based systems cannot, by allowing algorithms to map inputs to outputs without predefined rules.

Major tasks such as named entity recognition, machine translation, sentiment analysis, natural language understanding, and speech recognition are a result of the significant advancements NLP research has made. Emerging techniques such as graph-based methods and reinforcement learning combined with deep learning have advanced many NLP tasks to achieve or surpass human-level performance [12, p. 1].

1.2. LLMs and ChatGPT

1.2.1. what they are

Large Language Models have revolutionized the way we use computers to understand and process human language, driven by vast datasets and powerful neural architectures [13]. A Large Language Model is a type of an artificial intelligence system, which is able to interpret and generate human language [14]. The way LLMs work is based on the use of deep learning methods, resulting in the understanding and generation of accurate responses in countless subjects. Gan et al. [15] explain that LLMs are “deep learning models that learn the underlying patterns and rules of language by training on large-scale corpora”.

One of the most common comprehensive AI tools that is built on the top of LLMs is ChatGPT. ChatGPT is a revolutionary tool, which according to Wu et al. [16] is able to provide detailed responses according to an instruction in a prompt, whereas its capabilities can be proven in diverse natural language tasks, such as summarization, multilingual translation and question-answering [17]. It features billions of parameters and has been trained on a vast amount of text data including books and content from the web, making it a powerful instrument in the world of natural language processing. The history of ChatGPT showcases significant progress, with implications for various industries and advancements in the field of artificial intelligence [17].

1.2.2. ChatGPT's function

Kenney [18, p. 3] states that, besides being trained on 175 billion parameters in order to efficiently address user queries using predictive text, ChatGPT also utilizes reinforcement learning, supervised learning, and a 300-billion-word training dataset. Kenney also differentiates ChatGPT from other large language models architecture wise, stating that “While most large-scale language learning models use input to database matching algorithms, ChatGPT's approach relies instead on significant training data and human reinforcement to generate text in a predictive manner that (theoretically) allows it to answer questions quicker due to the reduction in computational processing power”

1.3. ChatGPT in education

1.3.1. education and technology

Making use of technology the right way can help us achieve an efficient and systematic learning method. According to Edinburgh [19], introducing technology to students can have a positive educational impact like the enhancement of learning and teaching, the improvement of educational outcomes and the development of digital skills. Technology's impact on education is profoundly transformative, revolutionizing how students learn. According to Kouser and Majid [20, p. 368] technology enriches students' learning experiences by offering the resources and tools they need to succeed, increasing their motivation.

1.3.2. the use of AI and ChatGPT by students

According to Arundel [21], based on a survey released by ACT¹, half of the students in high school actually make use of artificial intelligence for school and non-school tasks, with the most popular tool being ChatGPT. As stated by Forman et al. [22], one common reason why high school students use ChatGPT is the academic support, which helps improve their understanding of course content and enhances their time management skills. ChatGPT is not only able to generate responses, but also mimic human-like language making it an innovative, generative bot [23], in which students rely on for various projects.

1.4. ChatGPT in different subjects

1.4.1. ChatGPT in mathematics and physics

Considering that ChatGPT's existence has been solidified and that its use by high school students is inevitable, numerous studies have taken place to ascertain its abilities in generating accurate and reliable content in several subjects, such as mathematics, chemistry, the English language and many more.

Various research has taken place in order to validate ChatGPT's ability to solve mathematical tasks, with one of them being held by Dao and Le [24], who tested the model's capabilities in the following topics:

1. Combinations and Probability
2. Number Series (Arithmetic progression, Geometric progression)
3. Spatial Geometry
4. Exponential and Logarithmic Functions
5. Primitives and Integrals
6. Complex Numbers
7. Polyhedrons
8. Rotating Circle Block
9. Oxyz Spatial Calculus
10. Derivatives and Applications

¹ ACT is a nonprofit organization, that administers the college readiness exam (Online: <https://www.act.org/>, visited: 27.07.2024)

The questions were categorized into four levels of difficulty and were later on processed by ChatGPT to evaluate its accuracy and capabilities. The results highlight that ChatGPT performed well at the knowledge level, achieving 83% accuracy, but its performance declined significantly with higher difficulty levels, achieving only 10% accuracy at the high application level. The study found that ChatGPT excelled in topics like exponential and logarithmic functions but struggled with complex topics such as derivatives and spatial geometry. The limitations highlighted the need for improvement in handling more advanced mathematical concepts and higher-level problem-solving tasks, thereby pointing out the need for further investigation into its effectiveness and weaker points.

Rane [25] has also focused on the mathematical potential of the model, while also presenting an interesting table on his research, with a detailed explanation on the various roles and challenges that ChatGPT faces, while solving mathematical problems across different key-aspects, emphasizing the model's duality and proving, that challenges and limitations occur, showcasing its weaknesses.

Table 1: ChatGPT's mathematical capabilities according to Rane, N. [25]

Sr. No.	Mathematics Field	Role of ChatGPT	Challenges in Enhancement
1	Algebra	ChatGPT can provide assistance in solving algebraic equations, elucidating concepts, and offering step-by-step solutions.	Challenges include interpreting diverse problem formats and accommodating various user input styles.
2	Calculus	ChatGPT can aid in calculus problem-solving, differentiation, integration, and explaining intricate calculus concepts.	Challenges encompass managing complex symbolic expressions and generating accurate mathematical explanations comprehensible to users.
3	Geometry	ChatGPT can assist in solving geometric problems, explaining theorems, and	Challenges arise in accurately interpreting and generating geometric

		visualizing geometric shapes and their properties.	diagrams and logical proofs, ensuring precision.
4	Statistics	ChatGPT can support statistical analysis, explain statistical concepts, and interpret datasets, aiding in data-driven decision-making.	Challenges include managing extensive datasets, understanding context-specific statistical methods, and ensuring accurate interpretations to meet user needs.
5	Trigonometry	ChatGPT can assist in trigonometric problem-solving, explaining trigonometric functions, and solving various trigonometric equations.	Challenges involve handling a plethora of trigonometric identities and equations, ensuring precise and context-appropriate solutions.
6	Number Theory	ChatGPT can help with number theory concepts, prime factorization, and solving number-related problems and theorems.	Challenges involve managing large numbers, generating precise prime factorizations, and comprehending intricate number theory theorems.
7	Linear Algebra	ChatGPT can assist in solving systems of linear equations, performing matrix operations, and explaining abstract linear transformations.	Challenges include dealing with diverse matrix dimensions, ensuring accurate matrix computations, and explaining complex linear algebra concepts in an understandable manner.
8	Differential Equations	ChatGPT can aid in solving ordinary differential equations, explaining solution methods, and visualizing solutions in diverse scenarios.	Challenges encompass handling a variety of differential equation types, understanding boundary conditions, and ensuring the accuracy and applicability of solutions.

Note: Reprinted from "Enhancing Mathematical Capabilities through ChatGPT and Similar Generative Artificial Intelligence: Roles and Challenges in Solving Mathematical Problems," by N. Rane, 2023. SSRN. <http://dx.doi.org/10.2139/ssrn.4603237>

Besides mathematics, ChatGPT's abilities have been tested in more subjects and topics using different methods, tools, metrics and evaluations respectfully. The potential role of the model in physics education has been investigated by Liang et al. [26], specifically in solving calculation-based physics problems. The research evaluated ChatGPT's performance, focusing on its ability to handle numerical computations and understand physical concepts, such as vector variables and unit conversions. Through the use of prompting strategies to guide the model's reasoning step-by-step, the study demonstrated that ChatGPT can successfully solve certain physics problems, explain solutions, and generate new exercises. However, the model is not without flaws, often making errors due to incorrect assumptions about vector directions or misinterpretations of physical laws. These limitations highlight the importance of thoughtful integration of ChatGPT into educational settings.

1.4.2. ChatGPT and literature in education

However, studies around ChatGPT's abilities do not stop in hard sciences such as physics and chemistry. Researchers also focus on languages, literature, history, philosophy and many more, because ChatGPT "acts as a crucial research assistant, in the realm of Arts and Humanities" [27]. For example, Barad [28] provides on his paper a detailed presentation on ChatGPT and its use in the English literature classroom. More specifically, the experiment aimed to integrate AI into the teaching process to enhance creative writing, innovate instructional methods, and explore the tool's potential in literary analysis. ChatGPT was utilized to generate creative content such as poems inspired by "Waiting for Godot" and micro fiction based on the myth of Mansa Devi. Additionally, it assisted in creating teaching materials, including summaries, character analyses, and multiple-choice questions, and provided critical interpretations of literary texts. The experiment also examined how ChatGPT handled complex concepts like digital humanities and its responses to sensitive topics such as gender stereotypes, highlighting the need of recognizing possibly present biases in AI-generated content. The primary goals were to assess ChatGPT's effectiveness in supporting students' creative writing, making literature classes more interactive, and deepening literary analysis.

The outcomes were promising yet complex. ChatGPT successfully enriched creative writing and generated innovative teaching resources, making the learning experience more dynamic. However, it occasionally produced content that was culturally insensitive or lacked depth, underscoring the limitations of AI in grasping nuanced literary and cultural contexts. This experiment ultimately emphasized the need for educators to critically evaluate AI outputs, ensuring that while these tools can significantly enhance education, they are used thoughtfully and responsibly to maintain the quality and integrity of literary studies.

2. Empirical Section

The empirical section of this study is structured into four modules, each providing a detailed breakdown of the research process.

The first module focuses on data collection, offering a thorough examination of the data structure and the naming conventions applied to ensure clarity and consistency. It explains how the data is systematically organized and categorized to maintain uniformity across the dataset. Special attention is given to the patterns used during both the data collection and utilization stages, which play a key role in the overall data management strategy. Representative data samples are included to demonstrate the diversity of the dataset, and these are analyzed to identify underlying patterns and trends. In addition, various tables are provided, which outline the assigned IDs and key features, helping to evaluate and compare the generated responses in a structured manner.

The second module presents a sample “bundle” of information that has been converted into an XML file, highlighting the XML structure and its role in organizing the data. This module offers a detailed analysis of the XML tags, attributes, and the relationships between different components, showing how the structure supports the study’s research goals. Supporting tables are included to display key XML elements and their transformations, making it easier to understand how raw data is processed and converted into a standardized format. Visual representations, such as charts, are used to further illustrate the findings.

The third module focuses on a brief experiment with Meltemi, the first Greek LLM. The questions posed to this model are a selection of those asked to both ChatGPT and humans, with the same evaluation features applied to all three.

This section concludes with a summary of the study's results and final conclusions, offering a clear overview of the key insights gained from the research process.

2.1. **data collection and classification**

In order to organize the collected data, and later on convert them into individual XML files, a specific structure has been followed. More specifically, the information that has been grouped is the following:

1. A literary text
2. Corresponding question(s)
3. A given human answer
4. A generated ChatGPT answer
5. An instruction
6. An excel file

This group of information has been collected and formed in all cases and creates the so called "bundles". Every bundle consists of six files which later on have been used in order to create individual XML files for each case respectfully. In some cases, a bundle might consist of five files instead of six, because an instruction might not have been given at all. Each file has been named accordingly and follows the same pattern in all bundles. More specifically, here is how each file has been named after within the bundles:

1. The literary text → source_file_30056
2. The corresponding question(s) → text_file_question_30056
3. The given human answer → human_answer_30056
4. The generated ChatGPT answer → ChatGPT_answer_30056
5. An instruction → instruction_30056
6. An excel file → 30056

The number given right next to each file's name represents the code that the Greek Ministry of Education has given to each literary text respectfully and has been kept in this research for organizational purposes. Each literary text has been given an individual code.

As mentioned, this group of files creates a bundle. In the above exemplary case, the code 30056 corresponds to the bundle number “60” and consists of these six individual files, which have been enumerated right above. The total number of bundles is 232 and each one of them contains its unique corresponding files (either six or five). The following table displays a sample of the correlation between the codes given by the Ministry of Education and the corresponding codes used for enumerating the 232 formed bundles.

Table 2: Bundle IDs and Ministry IDs

Bundle IDs	Ministry’s IDs
1	23791
2	23793
3	23797
4	23798
5	23799
[...]	[...]
228	20147
229	20151
230	20152
231	20153
232	15396

2.2. bundle contents with examples

2.2.1. source file content

Having elucidated the structure of the bundles and the way that the individual files have been named and enumerated, it is crucial to highlight the information that each one of them contains, starting off with the literary text or else source file.

The source file text is divided into two sections. The first part provides an introduction, including the title and information on the historical context that the text was based on and/or on the author’s biography. The second part includes the literary text/poem itself.

First part:

Να ταξιδέψω;

Απόσπασμα από το ομότιτλο πεζό κείμενο του Πορτογάλου ποιητή και συγγραφέα Φερνάντο Πεσσόα (1888-1935) από «Το βιβλίο της Ανησυχίας» (εκδ. Αλεξάνδρεια, Αθήνα, 1997, μτφρ. Άν. Σπυράκου).

Second part:

Να ταξιδέψω; Για να ταξιδέψω φτάνει να υπάρχω: πηγαίνω από μέρα σε μέρα, σαν από σταθμό σε σταθμό στο σιδηρόδρομο του κορμιού μου ή του πεπρωμένου μου, σκυμμένος πάνω από τους δρόμους και τις πλατείες, πάνω από τα πρόσωπα και τις χειρονομίες, πάντα ίδια και πάντα διαφορετικά, όπως τελικά είναι και τα τοπία. Εάν φαντάζομαι, βλέπω. Τι παραπάνω κάνω ταξιδεύοντας; Μόνο μια αδυναμία ακραία της φαντασίας δικαιολογεί τη μετακίνηση σαν μέσο πλήρωσης των αισθήσεων. «Κάθε δρόμος, μέχρι κι αυτός ο δρόμος του Έντεπφουλ, θα σε οδηγήσει στην άκρη του κόσμου». Αλλά η άκρη του κόσμου, από τότε που ο κόσμος εξαντλήθηκε όταν τον φέραμε όλον βόλτα, είναι το ίδιο το Έντεπφουλ απ' όπου έφυγες. [...]

2.2.2. text file question content

The corresponding questions are located directly below the literary text. The number of these questions can range from one to three in each instance. In certain cases, a statement is provided just before the questions are formulated, serving as a conclusion derived from the literary text. In most cases students also receive short directions indicating for example the required word count for their answers or the elements they are expected to incorporate.

Στο Κείμενο 3 ο αφηγητής εκφράζει τη δική του οπτική για το ταξίδι. Ποια είναι αυτή και πώς αποτυπώνεται εκφραστικά; Να διατυπώσεις την προσωπική σου συμφωνία ή διαφωνία με αυτή την οπτική. Η απάντησή σου να εκτείνεται σε 150- 200 λέξεις.

Statement:

Στο Κείμενο 3 ο αφηγητής εκφράζει τη δική του οπτική για το ταξίδι.

First Question:

Ποια είναι αυτή και πώς αποτυπώνεται εκφραστικά;

Second Question:

Να διατυπώσεις την προσωπική σου συμφωνία ή διαφωνία με αυτή την οπτική.

Direction:

Η απάντησή σου να εκτείνεται σε 150-200 λέξεις.

Source: Τράπεζα Θεμάτων Διαβαθμισμένης Δυσκολίας, Ινστιτούτο Εκπαιδευτικής Πολιτικής, (<http://iep.edu.gr/el/trapeza-thematon-arxiki-selida>)²

2.2.3. human answer content

The number of answers corresponds to the number of questions dedicated to the literary text. For example, according to the literary questions formed above (First Question, Second Question), the human answer is expected to contain two equivalent parts (First Part, Second Part).

First part:

Για την οπτική του αφηγητή:

- 1) *δεν έχει νόημα να μετακινείται κάποιος, για να ταξιδέψει*
- 2) *αρκεί να παρατηρήσει τον εαυτό του ή τον περίγυρό του, για να νιώσει ότι ταξιδεύει, βιώνει δηλαδή νέες εμπειρίες*
- 3) *ο τόπος του καθενός εμπειριέχει την αρχική σύλληψη του κόσμου του ίδιου του ατόμου, στην οποία πάντα επιστρέφει*
- 4) *ο καθένας δημιουργεί τον κόσμο του, αρκεί να αξιοποιήσει τη φαντασία του*
- 5) *η ζωή είναι αυτό που εμείς την κάνουμε, επομένως τα ταξίδια είναι οι ίδιοι οι ταξιδιώτες*
- 6) *ό,τι βλέπουμε δεν είναι κάτι αντικειμενικό, αλλά αυτό που αντικρίζουμε μόνο μέσα από τη δική μας ματιά, άρα βλέπουμε αυτό που είμαστε.*

Εκφραστικά μέσα με τα οποία αποτυπώνεται η άποψη του αφηγητή:

- 1) *αρχικό ερώτημα, με το οποίο θέτει τον προβληματισμό του: Να ταξιδέψω;*

² Το/τα θέμα/τα προέρχεται και αντλήθηκε/αν από την πλατφόρμα της Τράπεζας Θεμάτων Διαβαθμισμένης Δυσκολίας που αναπτύχθηκε (MIS5070818-Τράπεζα θεμάτων Διαβαθμισμένης Δυσκολίας για τη Δευτεροβάθμια Εκπαίδευση, Γενικό Λύκειο-ΕΠΑΛ) και είναι διαδικτυακά στο δικτυακό τόπο του Ινστιτούτου Εκπαιδευτικής Πολιτικής (Ι.Ε.Π.) στη διεύθυνση (<http://iep.edu.gr/el/trapeza-thematon-arxiki-selida>)

2) ερωτήσεις, με τις οποίες θέλει να ενισχύει την κύρια τοποθέτησή του, ότι το ταξίδι είναι κάτι υποκειμενικό και δεν απαιτεί μετακίνηση: Τι παραπάνω κάνω ταξιδεύοντας; - Γιατί να ταξιδέψω;

3) μικροπερίοδος λόγος με αποφθεγματικού τύπου διατυπώσεις που προβληματίζουν τον αναγνώστη και τον βάζουν σε σκέψεις: Για να ταξιδέψω φτάνει να υπάρχω - Εάν φαντάζομαι, βλέπω - Μέσα μας είναι που τα τοπία έχουν τοπίο - Τα ταξίδια είναι οι ταξιδιώτες.

4) παρομοιώσεις: πηγαίνω... σαν από σταθμό σε σταθμό στο σιδηρόδρομο του κορμιού μου ή του πεπρωμένου μου

5) ασύνδετο σχήμα, με το οποίο αποδίδονται διάφοροι ταξιδιωτικοί προορισμοί σε όλο τον κόσμο: Στη Μαδρίτη, στο Βερολίνο, στην Περσία, στην Κίνα, στον καθένα από τους δύο Πόλους...

Second part:

Έκφραση συμφωνίας/διαφωνίας:

Η άποψη του αφηγητή με βρίσκει σύμφωνη, διότι υπογραμμίζεται η αξία του ρόλου της φαντασίας για κάθε άνθρωπο αναφορικά με τα ταξίδια. Αυτό που ο συγγραφέας επισημαίνει, είναι τα ταξίδια της ψυχής, που μπορεί να τον φτάσουν μέχρι την άκρη του κόσμου, όπως αυτός την ορίζει.

The human answer is a combination of the answer given by the Greek Ministry of Education and the answer given by me. That is because in most cases the second part of the answer is subjective, asking for example for one's opinion (agreement/disagreement) and is usually not given by the Ministry of Education. In order to fill in that gap the second part of the answer is most often completed based on my own perspective and expertise in the field. In some instances, the Ministry of Education has indeed provided an answer for all parts, leaving no answer incomplete.

2.2.4. ChatGPT's answer content

The next step after having collected and/or completed the expected human answer, is the collection of ChatGPT's response. More specifically, the prompt given to the model consisted of the literary text itself and its corresponding questions. Here is what ChatGPT's answer looks like based on the given prompt:

First part:

Ο αφηγητής αντιλαμβάνεται το ταξίδι ως μια εσωτερική εξερεύνηση, μια ευκαιρία να ανακαλύψει την πραγματικότητα του εαυτού του μέσα από τη διαίσθηση και τη φαντασία του. Βλέπει τη ζωή ως κατασκευή της προσωπικής μας πραγματικότητας και το ταξίδι ως μια απόδραση προς τον εαυτό μας, ανεξαρτήτως της φυσικής τοποθεσίας.

Second part:

Συμφωνώ με την οπτική αυτή επί του ταξιδιού. Κάθε ταξίδι, είτε είναι στον εξωτερικό κόσμο είτε στον εσωτερικό μας κόσμο, μπορεί να μας οδηγήσει σε αυτοσυνειδητοποίηση και ενδυνάμωση. Το να εξερευνούμε τις σκέψεις μας, τα συναισθήματά μας και τη σχέση μας με τον κόσμο μας είναι εξίσου σημαντικό με την εξωτερική περιπλάνηση. Τα ταξίδια μπορούν να είναι πηγή έμπνευσης και μάθησης, αλλά η πραγματική αλλαγή έρχεται όταν ανακαλύπτουμε τον εαυτό μας και την εσωτερική μας αλήθεια.

As seen on the generated outcome above, ChatGPT was able to capture both questions and respond accordingly. The model's response consists of two parts, which correspond to the questions formulated within the prompt. However, the content of the generated text and the quality of the answer are investigated and later on noted in the excel file.

2.2.5. excel file content

For the purpose of organizing and comparing the human answer with the one generated by ChatGPT, an excel file has been created. This file includes detailed information on the key features that should be addressed and analyzed within the answers. These features have been chosen based on the specific requests outlined in the questions and are expected to be highlighted and explained in the responses. The number of requested features may vary each time, but all of them have been referenced at some point within the provided questions. These features are essential for a complete response, as an answer may be deemed incomplete if they are not included. The table below (Table 3) outlines the selected features and provides corresponding examples chosen from some of the available questions.

Table 3: Expected features and corresponding examples

Features	Example
Χρήση στοιχείων/παραδειγμάτων μέσα από το κείμενο (use of elements/examples from the text)	Να ερμηνεύσεις τη στάση της γάτας, με στοιχεία του Κειμένου 2
Χρήση κειμενικών δεικτών (use of textual markers)	Να τεκμηριώσεις την απάντησή σου σχολιάζοντας τουλάχιστον τρεις (3) κειμενικούς δείκτες.
Χρήση εκφραστικών μέσων (use of figurative language)	Να ερμηνεύσεις την εικόνα του Παραδείσου στο Κείμενο 3 παραθέτοντας τρία εκφραστικά μέσα με τα οποία αποτυπώνεται στην αφήγηση
Έκφραση σκέψεων/συναισθημάτων - συμφωνίας/διαφωνίας (expression of thoughts/sentiments – agreement/disagreement)	Να διατυπώσεις την προσωπική σου συμφωνία ή διαφωνία με αυτή την οπτική
Αριθμός λέξεων (150-200) (word count)	Η απάντησή σου να εκτείνεται σε 150-200 λέξεις.

The use of elements/examples from the text is an essential feature, as it can substantiate a given answer, making it complete. This feature is required in all cases and its use can deem the answer as complete.

According to Kostopoulou [29] textual markers such as the literary style (ύφος) are designed to support an argument. These markers form a broader category, which also contains the element of figurative language.

Kostopoulou [30] describes figurative language (e.g. “hyperbole”) as a mean to persuade the audience, evoke emotions, and convey the author's opinions in writing. It is important to note that figurative language is a subcategory of the textual markers.

The expression of thoughts, sentiments, agreement and disagreement is an important part of the given answer and in most cases is also required. This aspect emphasizes subjectivity based on the literary text and the messages it conveys. The

students are able to express their personal beliefs and agree or disagree with the statements provided within the given literary text.

Lastly, the word limit is usually set between 150 and 200 words, though in some cases, it can be as low as 100 to 120 words. This range encourages the writer to focus on the essentials, offering a clear and direct answer without going off-topic or adding unnecessary details.

Besides the features mentioned above, it's important to disclose more details on how this information has been stored and compared. Figure 1 presents the contents of an exemplary excel sheet, which besides the features that have been analyzed on Table 3, includes additional information, which helps us understand the quality of the given answers.

Features	Source_1, Source_2	ChatGPT	Comments
Χρήση στοιχείων/παραδειγμάτων μέσα από το κείμενο	ναι	ναι	
Χρήση κειμενικών δεικτών	ΔΖ	ΔΖ	
			Human Αναφέρει παραπάνω από 3 εκφραστικά μέσα (6 σύνολο)
			ChatGPT Δεν αναφέρει ούτε ένα
Χρήση εκφραστικών μέσων	ναι	όχι	
Έκφραση σκέψεων/συναισθημάτων - συμφωνίας/διαφωνίας	ναι	ναι	
			Human: 200
Αριθμός λέξεων (150-200)	ναι	όχι	ChatGPT: 104

Figure 1: Excel file structure

The first column, as explained in Table 3, contains five essential features required for an answer to be considered complete. In the second and third columns, the following notations are used:

- "ναι", which stands for "yes" and indicates the existence of the expected feature.
- "όχι", which stands for "no" and indicates the lack of the expected feature.
- "ΔΖ", which stands for "not required", signifies that a feature is not expected/requested in this bundle.

As mentioned earlier, the second column includes the features shown within the human answer. Source_1 and source_2 have been chosen as notations, which indicate the answer's author. More specifically, source_1 refers to the answer given by the Ministry of Education, whereas source_2 refers to my own input. If an answer is entirely given by the Ministry of Education, column B will only display

source_1. If both notations appear, that means that the given human answer has been formed collaboratively.

The third column contains, as seen on Figure 1, the answers generated by ChatGPT. The model's performance has been marked accordingly either with a "yes", a "no" or a "ΔZ" based on its responses and the absence or occurrence of the expected key features.

The last column, column D, contains more detailed information, such as the number of features present in answers provided by ChatGPT and humans, the word count of each answer, indicators of any anomalies, and more. This data is essential for the comparative analysis between human and ChatGPT responses, allowing us to explore specifics and enabling data visualization at a later stage. At this stage it is important to point out the following:

1. An answer is deemed complete if all required features are present within the given answer. For example, if ChatGPT provides an answer, which consists of one or two correct textual markers rather than the required three, it is marked as "όχι" ("no") because the answer is not considered complete. This same rule applies to the corresponding human response.
2. For the required word count, typically set between 150 and 200 words, a margin of 10 words over or under is acceptable. Therefore, word counts of 140 to 210 words are considered within the acceptable range. For instance, on Figure 1 only the human answer is within the expected word count (human answer: 200 words, ChatGPT: 104 words)

After gathering the data, organizing it into bundles, categorizing the information within each excel file and comparing the features present in the human and ChatGPT answers, the next step consists of converting everything into XML files. XML is especially valuable in this case because it provides a structured, standardized way to organize and annotate data. According to Papaleo [31] "XML provides a basic syntax that can be used to share information between different applications and different organizations without using expensive and time-consuming conversion.", meaning that this converted dataset can be easily used for further purposes, such as training or finetuning a Large Language Model.

2.3. XML schema

```

<question id="60">
<min_id>30056</min_id>
<source>
  <intro>Να ταξιδέψω;
  Απόσπασμα από το ομότιτλο πεζό κείμενο του Πορτογάλου ποιητή και συγγραφέα Φερνάντο Πεσόα
  (1888-1935) από «Το βιβλίο της Ανησυχίας» (εκδ. Αλεξάνδρεια, Αθήνα, 1997, μεφρ. Αν.
  Σπυράκου).</intro>
  <text>Να ταξιδέψω; Για να ταξιδέψω φτάνει να υπάρχω: πηγαίνω από μέρα σε μέρα, σαν από σταθμό σε
  σταθμό στο σιδηρόδρομο του κορμιού μου ή του πεπρωμένου μου, σκυμμένος πάνω από τους δρόμους και τις
  πλατείες, πάνω από τα πρόσωπα και τις χειρονομίες, πάντα ίδια και πάντα διαφορετικά, όπως τελικά
  είναι και τα τοπία. Εάν φαντάζομαι, βλέπω. Τι παραπάνω κάνω ταξιδεύοντας; Μόνο μια αδυναμία ακραία
  της φαντασίας δικαιολογεί τη μετακίνηση σαν μέσο πλήρωσης των αισθήσεων. «Κάθε δρόμος, μέχρι κι
  αυτός ο δρόμος του Έντεπφουλ, θα σε οδηγήσει στην άκρη του κόσμου». Αλλά η άκρη του κόσμου, από τότε
  που ο κόσμος εξαντλήθηκε όταν τον φέραμε όλον βόλτα, είναι το ίδιο το Έντεπφουλ απ' όπου έφυγες.
  Στην πραγματικότητα η άκρη του κόσμου, όπως και η αρχή του, είναι η προσωπική μας σύλληψη του
  κόσμου. Μέσα μας είναι που τα τοπία έχουν τοπίο. Γι' αυτό, όταν τα φαντάζομαι, τα δημιουργώ. Αν τα
  δημιουργώ, υπάρχουν και, εφόσον υπάρχουν, τα βλέπω όπως βλέπω και τ' άλλα. Γιατί να ταξιδέψω; Στη
  Μαδρίτη, στο Βερολίνο, στην Περσία, στην Κίνα, στον καθένα από τους δύο Πόλους, πού αλλού θα
  βρισκόμουνα παρά μέσα σε μένα τον ίδιο, με τη δική μου ιδιαιτερότητα και το δικό μου τρόπο να
  αισθάνομαι. Η ζωή είναι αυτό που εμείς την κάνουμε να είναι. Τα ταξίδια είναι οι ταξιδιώτες. Αυτό
  που βλέπουμε δεν είναι αυτό που βλέπουμε, είναι αυτό που είμαστε.</text>
</source>
<questionPart>
  <statement>Στο Κείμενο 3 ο αφηγητής εκφράζει τη δική του οπτική για το ταξίδι.</statement>
  <subquestion id="1" direction="H απάντησή σου να εκτείνεται σε 150-200 λέξεις">Ποια είναι αυτή
  και πώς αποτυπώνεται εκφραστικά</subquestion>
  <subquestion id="2">Να διατυπώσεις την προσωπική σου συμφωνία ή διαφωνία με αυτή την
  οπτική</subquestion>
  <subquestion id="3"/>
  <subquestion id="4"/>
  <subquestion id="5"/>
</questionPart>
<instruction critical="False">
  <text>Οι μαθητές και οι μαθήτριες διατυπώνουν ελεύθερα τη συμφωνία ή τη διαφωνία τους με την
  οπτική του συγγραφέα, αρκεί να είναι η προσέγγισή τους σαφής και ολοκληρωμένη νοηματικά και
  εκφραστικά.</text>
</instruction>
<human_answers>
  <answer id="1">Για την οπτική του αφηγητή:
  1) δεν έχει νόημα να μετακινείται κάποιος, για να ταξιδέψει
  2) αρκεί να παρατηρήσει τον εαυτό του ή τον περίγυρό του, για να νιώσει ότι ταξιδεύει, βιώνει
  δηλαδή νέες εμπειρίες
  3) ο τόπος του καθενός εμπεριέχει την αρχική σύλληψη του κόσμου του ίδιου του ατόμου, στην οποία
  πάντα επιστρέφει
  4) ο καθένας δημιουργεί τον κόσμο του, αρκεί να αξιοποιήσει τη φαντασία του
  5) η ζωή είναι αυτό που εμείς την κάνουμε, επομένως τα ταξίδια είναι οι ίδιοι οι ταξιδιώτες
  6) ό,τι βλέπουμε δεν είναι κάτι αντικειμενικό, αλλά αυτό που αντικρίζουμε μόνο μέσα από τη δική μας
  ματιά, άρα βλέπουμε αυτό που είμαστε.
  Εκφραστικά μέσα με τα οποία αποτυπώνεται η άποψη του αφηγητή:
  1) αρχικό ερώτημα, με το οποίο θέτει τον προβληματισμό του: Να ταξιδέψω;
  2) ερωτήσεις, με τις οποίες θέλει να ενισχύει την κύρια τοποθέτησή του, ότι το ταξίδι είναι κάτι
  υποκειμενικό και δεν απαιτεί μετακίνηση: Τι παραπάνω κάνω ταξιδεύοντας; - Γιατί να ταξιδέψω;
  3) μικροπερίοδος λόγος με αποφθεγματικού τύπου διατυπώσεις που προβληματίζουν τον αναγνώστη και τον
  βάζουν σε σκέψεις: Για να ταξιδέψω φτάνει να υπάρχω - Εάν φαντάζομαι, βλέπω - Μέσα μας είναι που τα
  τοπία έχουν τοπίο - Τα ταξίδια είναι οι ταξιδιώτες.
  4) παρομοιώσεις: πηγαίνω... σαν από σταθμό σε σταθμό στο σιδηρόδρομο του κορμιού μου ή του
  πεπρωμένου μου
  5) ασύνδετο σχήμα, με το οποίο αποδίδονται διάφοροι ταξιδιωτικοί προορισμοί σε όλο τον κόσμο: Στη
  Μαδρίτη, στο Βερολίνο, στην Περσία, στην Κίνα, στον καθένα από τους δύο Πόλους...</answer>
  <answer id="2">Εκφραση συμφωνίας/διαφωνίας:
  Η άποψη του αφηγητή με βρίσκει σύμφωνη, διότι υπογραμμίζεται η αξία του ρόλου της φαντασίας για κάθε
  άνθρωπο αναφορικά με τα ταξίδια. Αυτό που ο συγγραφέας επισημαίνει, είναι τα ταξίδια της ψυχής, που
  μπορεί να τον φτάσουν μέχρι την άκρη του κόσμου, όπως αυτός την ορίζει.</answer>
  <answer id="3"/>
  <answer id="4"/>
  <answer id="5"/>
</human_answers>
<gpt_answers>
  <gpt_answer id="1">Ο αφηγητής αντιλαμβάνεται το ταξίδι ως μια εσωτερική εξερεύνηση, μια ευκαιρία
  να ανακαλύψει την πραγματικότητα του εαυτού του μέσα από τη διαίσθηση και τη φαντασία του. Βλέπει τη
  ζωή ως κατασκευή της προσωπικής μας πραγματικότητας και το ταξίδι ως μια απόδραση προς τον εαυτό
  μας, ανεξαρτήτως της φυσικής τοποθεσίας.</gpt_answer>
  <gpt_answer id="2">Συμφωνώ με την οπτική αυτή επί του ταξιδιού. Κάθε ταξίδι, είτε είναι στον
  εξωτερικό κόσμο είτε στον εσωτερικό μας κόσμο, μπορεί να μας οδηγήσει σε αυτοσυνειδητοποίηση και
  ενδυνάμωση. Το να εξερευνούμε τις σκέψεις μας, τα συναισθήματά μας και τη σχέση μας με τον κόσμο μας
  είναι εξίσου σημαντικό με την εξωτερική περιπλάνηση. Τα ταξίδια μπορούν να είναι πηγή έμπνευσης και
  μάθησης, αλλά η πραγματική αλλαγή έρχεται όταν ανακαλύπτουμε τον εαυτό μας και την εσωτερική μας
  αλήθεια.</gpt_answer>
  <gpt_answer id="3"/>
  <gpt_answer id="4"/>
  <gpt_answer id="5"/>
</gpt_answers>
</question>

```

Figure 2: Custom XML schema

Figure 2 showcases the XML schema of one out of the 232 bundles in total, which contains a plethora of information. Here is how the structure of the XML file was formed and what each element represents:

Table 4: Interpretation of the structure and the content of an XML file

Element	Interpretation	Example
<question id="60">	This code represents the number of the bundle (e.g. bundle No. 60). These codes are not officially given by the Ministry of Education; however, they have been used for organizational reasons. There are 232 bundles in total.	
<min_id>30056	This unique code has been assigned to a literary text and its corresponding question(s) by the Ministry of Education on its official website (https://www.trapeza-thematon.gr/Search). Each group of literary text and questions has its own code.	
<source>	This element represents the starting point of the literary text in the XML file.	
<intro>	Each literary text contains a short introduction. It provides information such as the title, the name of the author and/or some details about the historical background or the author's biography.	Να ταξιδέψω; Απόσπασμα από το ομότιπλο πεζό κείμενο του Πορτογάλου ποιητή και συγγραφέα Φερνάντο Πεσσόα (1888-1935) από «Το βιβλίο της Ανησυχίας» (εκδ. Αλεξάνδρεια, Αθήνα, 1997, μτφρ. Άν. Σπυράκου).
<text>	This element represents the appearance of the literary text itself.	Να ταξιδέψω; Για να ταξιδέψω φτάνει να υπάρχω: πηγαίνω από μέρα σε μέρα, σαν από σταθμό σε σταθμό στο σιδηρόδρομο του κορμιού μου ή του πεπρωμένου μου, σκυμμένος πάνω από τους δρόμους και τις πλατείες, πάνω από τα πρόσωπα και τις

		χειρονομίες, πάντα ίδια και πάντα διαφορετικά, όπως τελικά είναι και τα τοπία. Εάν φαντάζομαι, βλέπω. Τι παραπάνω κάνω ταξιδεύοντας; [...]
<questionPart>	This part sets the starting point for the corresponding questions to the text.	
<statement>	In some cases, a statement about the literary text is formulated right before its corresponding questions. This is not to be answered; however, it usually works as a short conclusion for the literary text.	Στο Κείμενο 3 ο αφηγητής εκφράζει τη δική του οπτική για το ταξίδι.
<subquestion id=>	Each literary text consists of corresponding questions that need to be answered by the students. The number of questions asked can vary in each case, usually between 1 and 3. As seen on Figure 2, there is a second subquestion for the case of bundle number 60.	<subquestion id="1"> Ποια είναι αυτή και πώς αποτυπώνεται εκφραστικά <subquestion id="2"> Να διατυπώσεις την προσωπική σου συμφωνία ή διαφωνία με αυτή την οπτική
direction	In some cases, right before or right after the literary questions are formed, a direction is given, which works as a guideline and should be followed by the students, in order for their answer to be considered correct. The position of the direction within the XML files is always right next to the first subquestion, as seen on Figure 2.	Η απάντησή σου να εκτείνεται σε 150-200 λέξεις

<p><instruction critical = ></p>	<p>This specific type of instruction is found within the correction sheet of the person responsible for evaluating the students' responses and it works as a guideline. The instruction is divided into "critical" and "not critical". Within the XML files this division is set as follows:</p> <ol style="list-style-type: none"> 1. <instruction critical = False> 2. <instruction critical = True> <p>An instruction is considered critical when it provides detailed and unique information tailored to a specific literary text. On the other hand, a non-critical instruction is considered generic and recycled on many different bundles, making little to no difference to the evaluator's work.</p>	<p>Οι μαθητές και οι μαθήτριες διατυπώνουν ελεύθερα τη συμφωνία ή τη διαφωνία τους με την οπτική του συγγραφέα, αρκεί να είναι η προσέγγισή τους σαφής και ολοκληρωμένη νοηματικά και εκφραστικά.</p>
<p><human_answers></p>	<p>Human answers work as the ground truth. Each human answer is structured based on the literary text's corresponding questions. For example, if the questions are three in total, so will the human answers.</p> <ol style="list-style-type: none"> 1. <answer id="1"> 2. <answer id="2"> 3. <answer id="3"> 4. <answer id="4"> 5. <answer id="5"> <p>As seen on Figure 2 the questions formulated based on the literary text are two. Based on this fact we</p>	<p><answer id="1"> Για την οπτική του αφηγητή: 1) δεν έχει νόημα να μετακινείται κάποιος, για να ταξιδέψει 2) αρκεί να παρατηρήσει τον εαυτό του ή τον περίγυρό του, για να νιώσει ότι ταξιδεύει, βιώνει δηλαδή νέες εμπειρίες [...]</p> <p><answer id="2"> Έκφραση συμφωνίας/διαφωνίας: Η άποψη του αφηγητή με βρίσκει σύμφωνη, διότι υπογραμμίζεται η αξία του ρόλου της φαντασίας για κάθε άνθρωπο αναφορικά με τα ταξίδια. Αυτό που ο συγγραφέας επισημαίνει, είναι τα ταξίδια της</p>

	can observe that the number of given human answers aligns with the number of questions asked.	ψυχής, που μπορεί να τον φτάσουν μέχρι την άκρη του κόσμου, όπως αυτός την ορίζει.
<gpt_answers>	<p>In order for ChatGPT to generate a text, it is given the exact same literary text and its assigned questions. Then the model responds to the questions accordingly formulating its own generated answers. In many cases the model is able to respond to all questions asked. However, it is possible, that it won't generate an answer for each question always. This has actually been spotted in some of the 232 cases. The structure of the GPT answers within the XML looks like the one used for the human answers:</p> <ol style="list-style-type: none"> 1. <gpt_answer id="1"> 2. <gpt_answer id="2"> 3. <gpt_answer id="3"> 4. <gpt_answer id="4"> 5. <gpt_answer id="5"> <p>In the case used as an example (bundle 60) the model was able to respond to both questions accordingly.</p>	<p><gpt_answer id="1"> Ο αφηγητής αντιλαμβάνεται το ταξίδι ως μια εσωτερική εξερεύνηση, μια ευκαιρία να ανακαλύψει την πραγματικότητα του εαυτού του μέσα από τη διαίσθηση και τη φαντασία του. Βλέπει τη ζωή ως κατασκευή της προσωπικής μας πραγματικότητας και το ταξίδι ως μια απόδραση προς τον εαυτό μας, ανεξαρτήτως της φυσικής τοποθεσίας.</p> <p><gpt_answer id="2"> Συμφωνώ με την οπτική αυτή επί του ταξιδιού. Κάθε ταξίδι, είτε είναι στον εξωτερικό κόσμο είτε στον εσωτερικό μας κόσμο, μπορεί να μας οδηγήσει σε αυτοσυνειδητοποίηση και ενδυνάμωση. Το να εξερευνούμε τις σκέψεις μας, τα συναισθήματά μας και τη σχέση μας με τον κόσμο μας είναι εξίσου σημαντικό με την εξωτερική περιπλάνηση [...]</p>

3. Meltemi

3.1. **introduction to Meltemi**

Meltemi³ is the first open-source LLM for the Greek language trained by the Institute for Language and Speech Processing of the Athena Research and Innovation Center⁴. With 7 billion parameters, Meltemi was trained on a massive 40-billion-token corpus of Greek language data to capture and generate Greek language nuances accurately. The project used continual pretraining based on the Mistral architecture, optimized with both Greek monolingual and Greek-English bilingual datasets, enhancing the model's ability to handle language-specific tasks and provide improved translations and natural language understanding for Greek. In addition to the main model, the developers created Meltemi 7B Instruct, a chat-focused version fine-tuned for interactive applications. This version is aligned with human preferences and has undergone careful filtering to remove toxic content, making it suitable for public-facing uses. Meltemi 7B Instruct is particularly aimed at improving accessibility to Greek-language AI solutions in sectors such as customer service, digital communication, and language preservation.

The creation of Meltemi responds to the need for advanced AI models tailored to underrepresented languages like Greek compared to widely supported ones. The model was created by Athena Research Center with the goal of enabling local AI applications with a profound comprehension of Greek linguistic and cultural traits. This model fosters more approachable and pertinent AI-driven solutions in Greek by offering a useful resource for creating, analyzing and translating content in the Greek language [32].

3.2. **experimenting with Meltemi**

To evaluate Meltemi's capabilities, 10 of the 232 available literary questions were selected. The same question-asking process used for ChatGPT was followed, although the data classification process differed, as it was not divided into bundles and converted into individual XML files. The goal of this brief experiment is to have a glimpse of the model's abilities when it comes to literary analysis. The

³ <https://meltemi.ilsp.gr/>

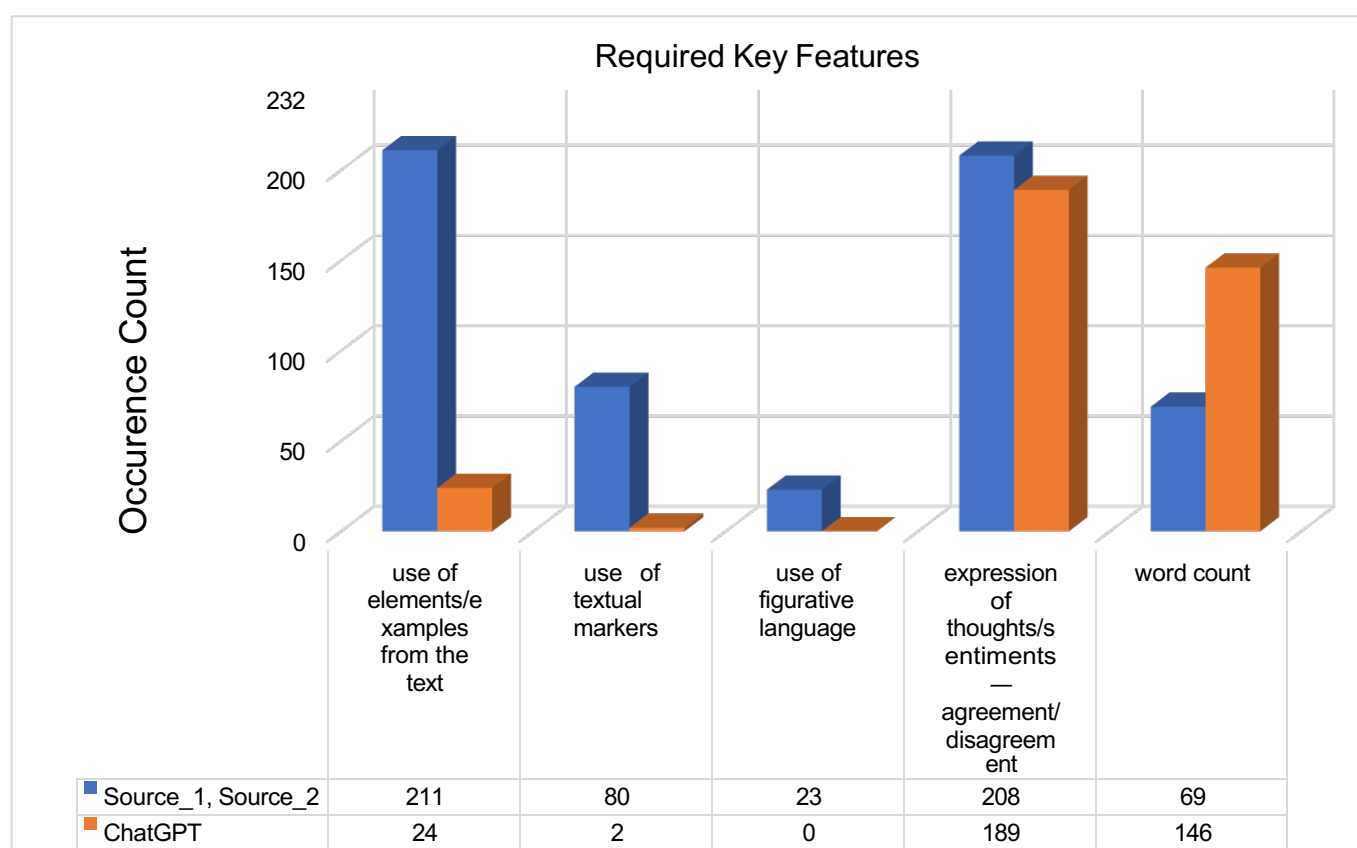
⁴ Athena Research Center: <https://www.athenarc.gr/en/home> (visited: 10.11.2024)

structure of the given literary texts and their corresponding literary questions are mentioned in Chapter 2 and more specifically in the subcategories [2.2.1](#) and [2.2.2](#).

4. Results

4.1.1. occurrence count of required key features (human, ChatGPT)

The following chart highlights the number of times that the required five features were successfully mentioned as expected within the total 232 given human and ChatGPT answers.



Graph 1: Comparison of Key Feature Occurrences in Human and ChatGPT Responses

The numbers in the chart indicate the times each feature was present in the responses, with a maximum possible count of 232.

Use of elements/examples from the text is far more common in human responses, with 211 occurrences compared to ChatGPT's 24. This suggests that human responses often draw directly from the literary texts, using specific examples or elements, while ChatGPT does so much less frequently.

For use of textual markers human responses again show a higher frequency, with 80 occurrences, whereas ChatGPT responses include only 2. This difference implies that humans more naturally employ markers to improve the flow and coherence of their answers, while following the given instructions given by the Ministry of Education.

Figurative language is rare in both types of responses, though it appears exclusively in human answers, with 23 instances, and none in ChatGPT responses. This indicates that the use of figurative expressions is uncommon overall but may be particularly challenging or less typical for ChatGPT.

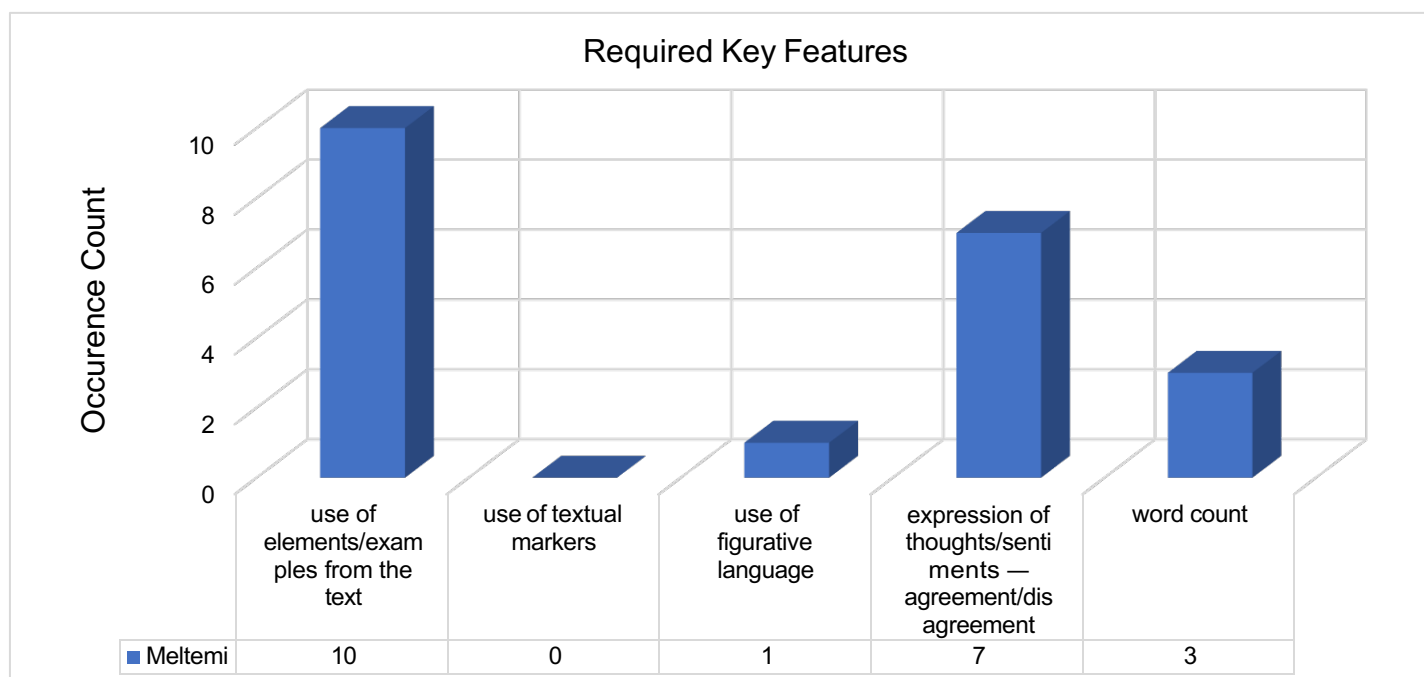
In both cases of ChatGPT's low performance in spotting textual markers and figurative language within the literary texts, it is important to note that the model's responses are often considered incomplete or irrelevant. The model has indeed been able to locate such elements in various instances, as seen in [4.3.](#), however many of its responses have been marked as not acceptable, because they are considered incomplete.

In terms of expression of thoughts/sentiments - agreement or disagreement, both human and ChatGPT responses frequently reflect these elements, with humans showing 208 occurrences and ChatGPT 189. This similarity suggests that ChatGPT is fairly adept at conveying opinions, demonstrating a level of subjective engagement comparable to humans.

Word count is notably higher in ChatGPT responses, totaling 146 occurrences compared to 69 in human responses. This suggests that ChatGPT tends to be more consistent to staying within the expected word count.

4.1.2. **occurrence count of required key features (Meltemi)**

The elements evaluated in Meltemi's answers pertain to the same key features that were examined in the responses provided by both ChatGPT and humans. Graph 2 displays the numbers in the chart indicate the times each feature was present in the responses, with a maximum possible count of 10.



Graph 2: Key Feature Occurrences in Meltemi's Responses

As shown in Graph 2, Meltemi's capabilities can vary based on the requested key feature.

The model consistently utilized textual elements and examples to support its responses in all ten instances.

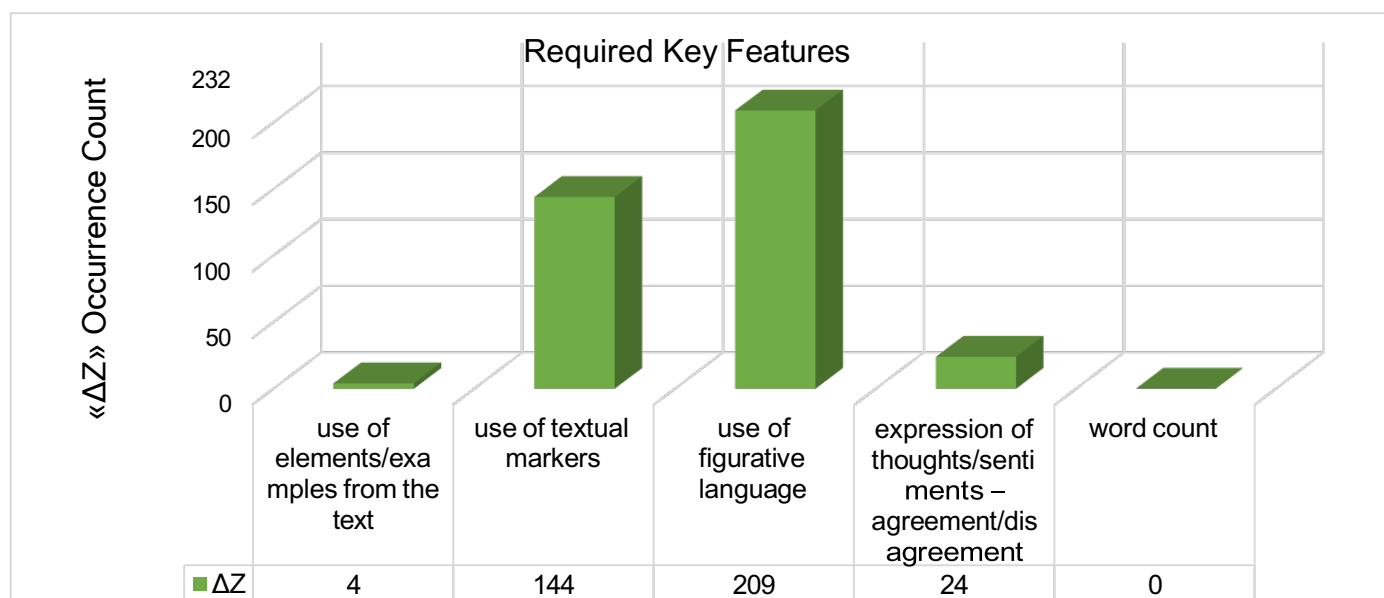
However, the model struggled to identify and employ three or more textual markers or figurative language means. While it managed to identify one or two in certain cases, as seen in [4.4](#), it fell short of the required threshold. As mentioned earlier, some responses may have been marked as not acceptable and they might have been considered incomplete, in case the model has not been able to locate at least three of the requested textual markers/figurative language means.

The number of times the model has expressed a personal thought/sentiment — agreement/disagreement on a requested literary text is quite high, achieving success in seven out of ten instances.

Lastly, Meltemi struggled to adjust to word count constraints, exceeding the limit in 70% of cases, highlighting an inconsistency in maintaining textual brevity.

4.2. occurrence count of “ΔZ” in required key features

Graph 3 displays the occurrence of “ΔZ” (“not required”) within all excel files. It is important to highlight, that the existence of all required key features simultaneously is not mandatory and this is why “ΔZ” has been used as a corresponding notation.



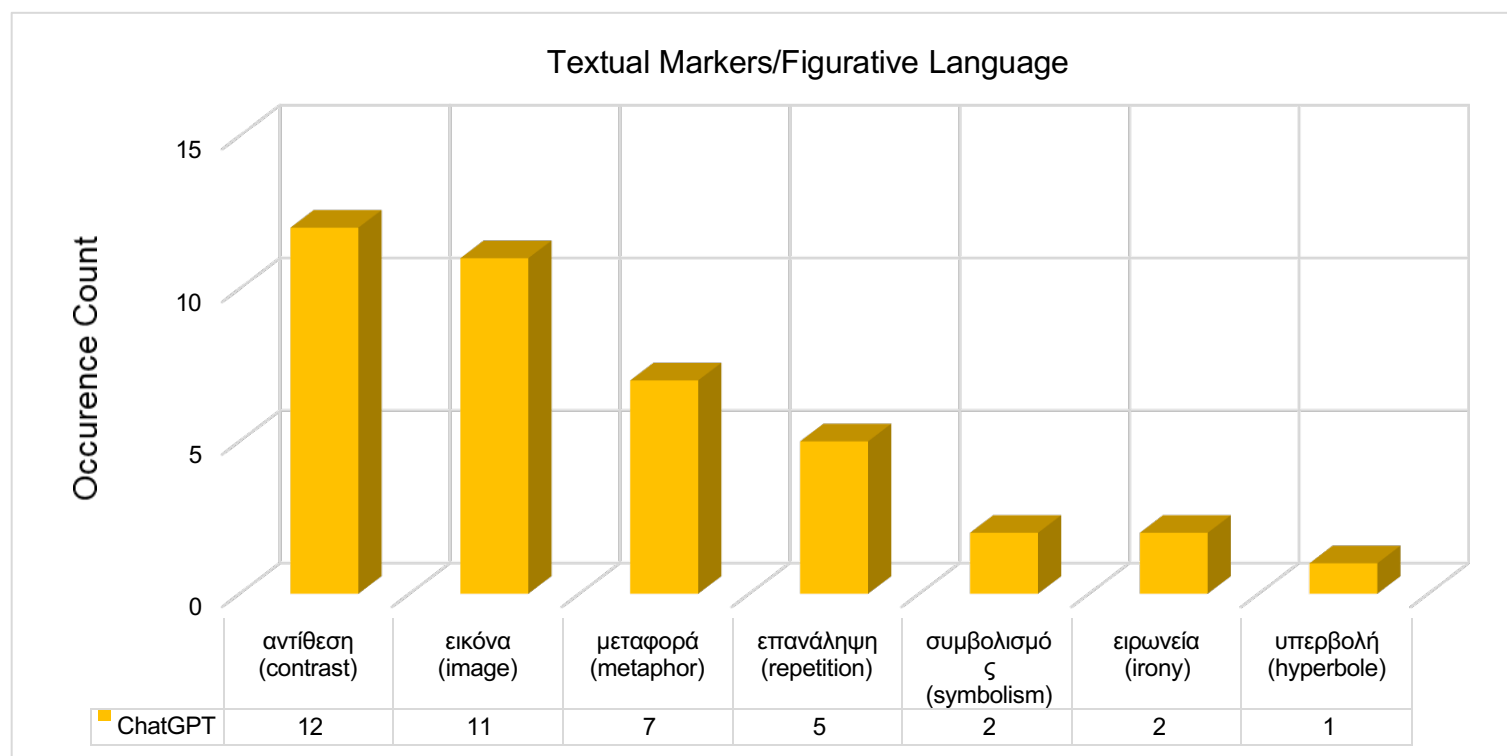
Graph 3: Frequency of Non-Required Features in Answers (Marked with “ΔZ”)

As shown in Graph 3, figurative language is the least frequently required feature, while word count is always included in the questions that support the literary texts.

Emphasizing the presence of “ΔZ” is crucial to clearly differentiate instances where these five features are either present or absent in the responses. Solely displaying Graph 1, which only highlights the occurrence of these five features, could lead to misinterpretation, suggesting that other responses simply lack required features, whether generated by humans or ChatGPT. However, as shown in Graph 3, “ΔZ” specifically denotes cases where a feature is not expected to appear in the answers, providing a more accurate understanding.

4.3. ChatGPT's most referenced textual markers/figurative language

As previously noted, ChatGPT has limited ability to accurately identify textual markers and figurative language within the literary text. The model's responses often contain either too few elements or elements that do not qualify as such. The following graph highlights the textual markers and figurative language means that ChatGPT correctly identified and referenced most frequently.



Graph 4: ChatGPT's most referenced textual markers & figurative language

As shown in Graph 4, ChatGPT successfully identified a total of forty accurate textual markers and figurative language means, with “contrast” emerging as the most frequent, appearing in twelve instances. Additional features such as “image” and “metaphor” are also detected within ChatGPT's responses, while elements like “irony” and “symbolism” appear with lesser frequency.

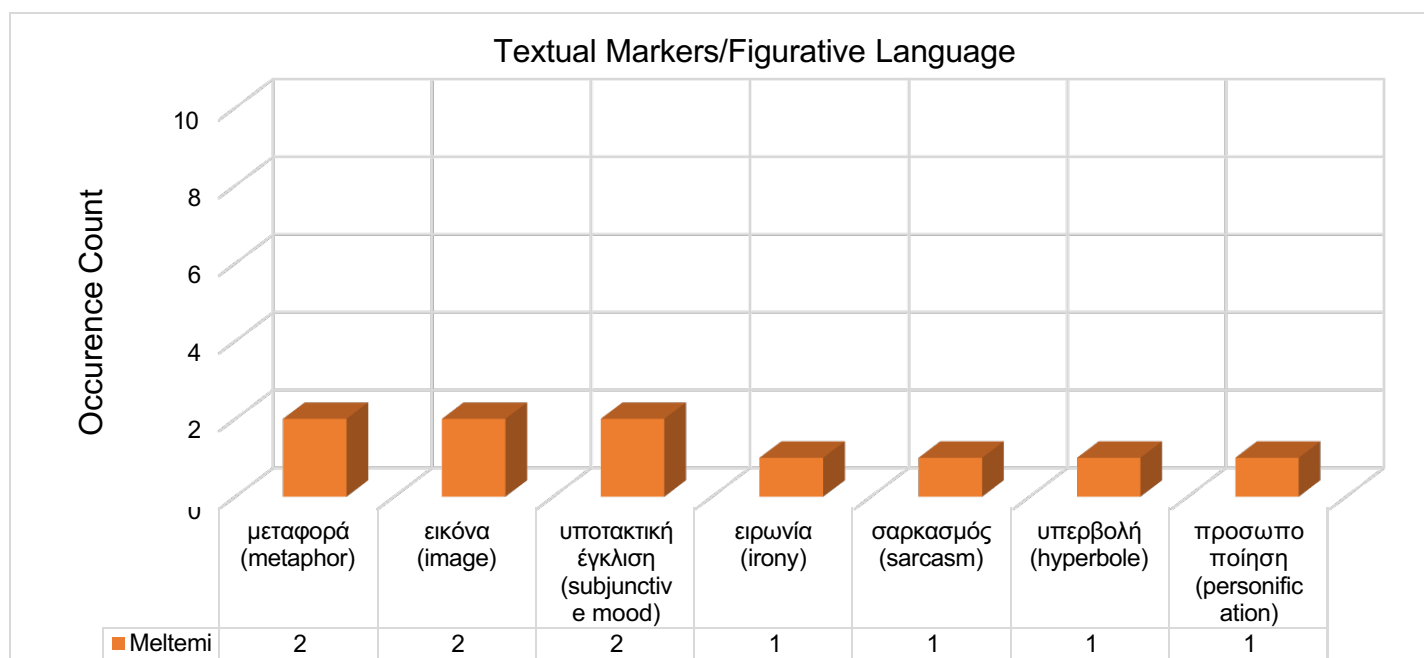
Despite the model's ability to identify certain features, its responses have been deemed incorrect in most cases, as indicated by the “όχι” label in Figure 1. This designation reflects that the responses were considered incomplete, either because the model failed to provide at least three textual markers/figurative language means as mentioned within the literary question or because the ones it

actually provided were inaccurate. For example, in bundle “23” ChatGPT has marked the following as textual markers within its answer:

0. “κούραση” (exhaustion)
1. “πόθος” (desire)
2. “απόγνωση” (desperation)

These elements are not considered textual markers and thus ChatGPT’s answer has been marked inaccurate and notated with an “όχι” as explained earlier.

4.4. Meltemi’s most referenced textual markers/figurative language



Graph 5: Meltemi’s most referenced textual markers & figurative language

As shown in Graph 5, Meltemi successfully managed to identify a total of ten accurate textual markers and figurative language means, with “metaphor”, “image” and the “subjunctive mood” emerging as the most frequent ones, appearing in two instances each. Additionally, the model detected instances of “irony” and “personification” within some of the given literary texts.

Meltemi’s ability to stay within the requested word count (either 120-150 or 150-200, depending on the case) is unfortunately low, as the model was able to maintain its answers within that frame only 30% of the times, indicating its incapability to deliver a completely acceptable response.

5. Conclusion

This study examined the capabilities of ChatGPT, a leading Large Language Model, in literary analysis while directly comparing its performance to human-generated responses from the Greek Ministry of Education and the author. By analyzing 232 literary texts from the Greek high school curriculum, the research sought to evaluate ChatGPT's ability to respond to literary questions while highlighting its strengths and weaknesses against human responses. The analysis was based on five key literary features: the use of examples/elements from the text, textual markers, figurative language, subjective agreement or disagreement, and adherence to word count.

The findings underscore ChatGPT's strengths as well as its limitations in this context. The model excelled in subjective expression, often demonstrating a strong ability to articulate agreement or disagreement with literary perspectives. Its adherence to word count guidelines was also notable, offering responses that generally stayed within the required range. However, it struggled with identifying and applying more nuanced literary features, such as textual markers and figurative language. These shortcomings resulted in incomplete or less contextually accurate responses when compared to human-generated answers, which consistently demonstrated depth, contextual understanding, and alignment with the evaluative criteria.

Furthermore, the experiment with Meltemi, the first Greek open-source LLM, provided additional insights into the potential of language models developed specifically for Greek. While Meltemi demonstrated an ability to integrate textual elements effectively, its performance on the number of required literary features and word count constraints highlighted areas for improvement. The findings emphasize the importance of continued development of Greek-specific AI tools to meet the unique linguistic and educational challenges of the language.

The conversion of the bundles into XML files added significant value to the study, creating a structured dataset that is not only suitable for the analysis conducted for this research, but also for future applications in training or fine-tuning LLMs tailored to the Greek language. This dataset represents a step forward in bridging the gap between educational needs and AI capabilities, offering a scalable

framework for further research in natural language processing for underrepresented languages like Greek.

Overall, this thesis demonstrates the value of integrating AI into educational settings while underscoring the necessity of human oversight and intervention. ChatGPT's responses, though promising, reveal inherent limitations when tasked with subjective, contextually rich analyses like literary evaluation. By providing a comparative framework and structured data, this research contributes to the evolving discussion surrounding AI in education, particularly for the Greek language. It encourages the development of more targeted AI models and the adoption of methodologies that combine human expertise with machine efficiency to enhance learning outcomes.

Terminology Table

Greek terminology	English terminology
κειμενικοί δείκτες	textual markers
εκφραστικά μέσα	figurative language
αριθμός λέξεων	word count
ναι	yes
όχι	no
αντίθεση	contrast
μεταφορά	metaphor
εικόνα	image
επανάληψη	repetition
συμβολισμός	symbolism
ειρωνεία	irony
υπερβολή	hyperbole
υποτακτική έγκλιση	subjunctive mood
προσωποποίηση	personification
κούραση	exhaustion
πόθος	desire
απόγνωση	desperation

Abbreviations – Acronyms – Initialisms

NLP	Natural Language Processing
LLM	Large Language Model
XLM	Extensible Markup Language
ChatGPT	Chat Generative Pre-trained Transformers
AI	Artificial Intelligence
ΔΖ	Δεν Ζητείται

References

- [1] Dash, S., & Bhoi, C. (2024). Exploring the intersection of education and artificial intelligence: A comprehensive review. *International Journal of Multidisciplinary Approach Research and Science*, 2(2), 601–610. <https://doi.org/10.59653/ijmars.v2i02.637>
- [2] Azaria, A. Azoulay, R., & Reches, S. (2024). ChatGPT is a remarkable tool-for experts. *Data Intelligence*, 6(1), 240-296. https://doi.org/10.1162/dint_a_00235
- [3] Saussure, F. de. (1916). *Course in general linguistics* (C. Bally & A. Sechehaye, Eds.; W. Baskin, Trans., p. 9). McGraw-Hill. (Original work published 1916)
- [4] Chomsky, N. (2002). *On nature and language*. Cambridge University Press.
- [5] Sweet, H. (1900). *The history of language*. J.M. Dent.
- [6] Grenoble, L., & Whaley, L. (2006). *Saving languages: An introduction to language revitalization*. Cambridge University Press.
- [7] Oduor, S. O., & Wang, J. (2019). *Science & technology revolution and human language*. Proceedings of the 2nd International Conference on Contemporary Education and Economic Development (CEED 2019), Wuhan University, China. CSP.
- [8] Del Vecchio Lança, L. (2022). *Humans, language, and technology: The interplays between language and technology according to Heidegger's philosophy* (Final degree project, Universitat de Vic – Universitat Oberta de Catalunya). Retrieved from https://openaccess.uoc.edu/bitstream/10609/145824/6/ldelvecchioFBDP0622r_eport.pdf
- [9] Bardeen, R., & Cerpa, S. (2015). Language and technological development: A growing relationship. *Journal of Social Science and Technology*, 11(3), 221– 233.
- [10] Gallagher, S., Rafferty, A., & Wu, A. (2004). *Natural language processing: Overview and history*. Stanford University. Retrieved from https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview_history.html
- [11] Johri, P., Khatri, S. K., Al-Taani, A., Sabharwal, M., Suvanov, S., & Chauhan, A. (2021). Natural language processing: History, evolution, application, and future work. In *Advances in Science, Technology & Innovation* (pp. 369-382). Springer. https://doi.org/10.1007/978-981-15-9712-1_31
- [12] Montejo-Ráez, A., & Jiménez-Zafra, S. M. (2022). Current approaches and applications in natural language processing. *Applied Sciences*, 12(10), 4859. <https://doi.org/10.3390/app12104859>
- [13] Karanikolas, N., Manga, E., Samaridi, N., Tousidou, E., & Vassilakopoulos, M. (2023). Large Language Models versus Natural Language Understanding and Generation. In *PCI 2023: 27th Pan-Hellenic Conference on Progress in Computing and Informatics*, Lamia, Greece, November 24-26, 2023.

[14] Törnberg, P. (2023). How to use large language models for text analysis. Institute of Language, Logic and Computation (ILLC), University of Amsterdam.

[15] Gan, W., Qi, Z., Wu, J., & Lin, J. C. W. (2023). Large language models in education: Vision and opportunities. *Proceedings of the IEEE International Conference on Big Data*, 1–10. <https://doi.org/10.48550/arXiv.2311.13160>

[16] Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5), 1023-1038. <https://doi.org/10.1109/JAS.2023.1000571>

[17] Gupta, N. K., Chaudhary, A., Singh, R., & Singh, R. (2023). ChatGPT: Exploring the capabilities and limitations of a large language model for conversational AI. 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), Faridabad, India, pp. 139-142. <https://doi.org/10.1109/ICAICCIT60255.2023.10465811>

[18] Kenney, N. (2023). A brief analysis of the architecture, limitations, and impacts of ChatGPT. Zenodo. <https://doi.org/10.5281/zenodo.7762245>

[19] Edinburgh (2016). *Enhancing Learning and Teaching Through The Use Of Digital Technology*. UNESCO. https://planipolis.iiep.unesco.org/sites/planipolis/files/ressources/scotland_dig_tech_00505855.pdf

[20] Kouser, S., & Majid, I. (2024). Technological tools for enhancing the teaching and learning process. *Teaching & Education*, 13(1), 366-373. <https://doi.org/10.37867/TE130133>

[21] Arundel, K. (2023). How are high schoolers using AI? K-12 Dive. <https://www.k12dive.com/news/artificial-intelligence-high-school-students/702228/>

[22] Forman, N., Udvaros, J., & Avornicului, M. (2023). ChatGPT: A new study tool shaping the future for high school students. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(1), 95-102. <https://doi.org/10.59287/ijanser.562>

[23] Lock, S. (2022). What is AI chatbot phenomenon ChatGPT and could it replace humans? *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans>

[24] Dao, X.-Q., & Le, N.-B. (2023). Investigating the effectiveness of ChatGPT in mathematical reasoning and problem-solving: Evidence from the Vietnamese National High School Graduation Examination. arXiv preprint arXiv:2306.06331. <https://arxiv.org/abs/2306.06331>

[25] Rane, N. (2023). Enhancing mathematical capabilities through ChatGPT and similar generative artificial intelligence: Roles and challenges in solving mathematical problems. <http://dx.doi.org/10.2139/ssrn.4603237>

- [26] Liang, Y., Zou, D., Xie, H., & Wang, F. L. (2023). Exploring the potential of using ChatGPT in physics education. *Smart Learning Environments*, 10(1), 1-14. <https://doi.org/10.1186/s40561-023-00273-7>
- [27] Rane, N. (2023). Role and challenges of ChatGPT and similar generative artificial intelligence in arts and humanities. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4603208>
- [28] Barad, D. (2023). *Exploring ChatGPT in the English Literature Classroom*. <https://doi.org/10.13140/RG.2.2.32590.02889>
- [29] Kostopoulou, M. (n.d.). Κειμενικοί Δείκτες. *Filol.* Retrieved [10.11.2024], from <https://filol.gr/thewria/>
- [30] Kostopoulou, M. (n.d.). *Γλωσσικές Επιλογές ή Εκφραστικά μέσα*. *Filol.* Retrieved from <https://filol.gr/thewria/>
- [31] Papaleo, L. (2013). *Introduction to XML and its applications*. https://doi.org/10.1142/9789812836304_0006
- [32] Voukoutis, L., Roussis, D., Paraskevoopoulos, G., Sofianopoulos, S., Prokopidis, P., Papavasileiou, V., Katsamanis, A., Piperidis, S., & Katsouros, V. (2024). *Meltemi: The first open Large Language Model for Greek*. Athena Research Center. Retrieved from <http://arxiv.org/abs/2407.20743>