N° d'ordre :
N° de série :

الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جامعة الشهيد حمه لخضر الوادي
University of Echahid Hamma Lakhdar El-Oued
كلية العلوم الدقيقة
Faculty of Exact Sciences
قسم الإعلام الآلي
Computer Science Department

Thesis
Submitted in partial fulfilment of the requirements for the Degree of

ACADEMIC MASTER

Field: Mathematics and Computer Science
Option: Computer Science
Speciality: Distributed Systems and Artificial Intelligence

Theme

Linguistic Inquiry Word Count and Deep Learning for Multimedia Analysis

Presented by:

- SAYADI Rabia
- SELMI Amina

Discussed on June 2, 2024 by the jury:

NAOUI Mohammed Anouar	PLB	Supervisor	Univ. El Oued
MEFTAH Mohammed Charef Eddine	Professor	President	Univ. El Oued
GHERBI Kaddour	APA	Examiner	Univ. El Oued

Academic year: 2023/2024

Dedication

In the name of Allah, the Most Gracious, the Most Merciful, All praise and thanks are due to Allah for his endless blessings and guidance. Without His grace, this work would not have been possible.

To our beloved parents, who have been our source of inspiration, guide, and give us strength when we thought of giving up, who continually provide their moral, spiritual, emotional, and financial support.

*To our research supervisor, '**Naoui Mohammed Anouar**' for his helping, and guiding us, thank you.*

To our families, brothers and sisters:
For their support and patient.

To our friends who became our supporters and helped us to any problem we faced. They encourage us to finish this important research on time. They also give us inspirational messages every time.

To our teachers who believed that we would finish this research in time, and inspiring us with their inspirational stories when they were students before.

Dedication

Whoever said, “I am hers”, “got her.” And I was hers. The journey was not short, nor was the road fraught with ease, but I did it. Praise be to Allah, who facilitated the beginnings and made us reach the endings.

To my ambitious, great, strong self first, who endured all the pitfalls and continued despite the difficulties. It began with ambition and ended with success, then to everyone.

To the one who worked with me to complete my university career, to the one whose name I carry with pride, to the one who paved the way for us with effort, and to the woman who made me an ambitious girl who loves challenges and difficulties, my first role model, to those who showered us with love and bore with us the hardships of the path, to those whom I promised this success, I have fulfilled my promise, and I give it to you dad and mom.

To the shoulder that does not lean and the shadow in which I take refuge. To the hearts beating with sincerity of love and feelings. To the stable pillars in life.

To my brothers.

To those who believed in me and my abilities. To the hearts beating with sincerity of love and feelings. To the ones I can rely on no matter what. Your love and guidance mean the world to me.

To my sisters.

To the pure hands that removed the thorns of failure from my path.

To you, my family.

To my dear friend **Selmi Amina***, for her steadfast support and understanding over the years.*

To my lovely friend, **Gouni Ahlem** *for believing in me and providing unwavering confidence. Your friendship is invaluable to me.*

To those who bet on my arguments and remind me of how strong and capable I am. Comrades over the years, the journey was not short and things were not easy, but... With God's help, I did it.

I dedicate to you all this humble work and the fruit of my effort.

God grants success.

Sayadi Rabia

Dedication

In the name of Allah, the Most Gracious, the Most Merciful,
All praise and thanks are due to Allah for His endless blessings and guidance. Without His grace, this work would not have been possible.

To myself, for pushing through, for perseverance and hard work that made this journey possible, and for finishing what I started.

To my parents, who gave me the gift of dreams and the ability to realize them.

To my dear sisters and brothers, Thank you for your unwavering support, endless patience, and boundless love.

To everyone who taught me, guided me, and inspired me throughout my academic journey.

To those rare, beautiful friendships that changed my life forever.

To my dear friend ***Sayadi Rabiaa****, for her camaraderie, understanding, and support through the years.*

To my lovely teacher and best friend, ***Bessei Tourquia****, for believing in me and being a source of unwavering confidence. Your friendship means the world to me.*

Thank you all for believing in me and for being part of this journey.

Selmi Amina

Abstract

This work is devoted to the integration of Linguistic Inquiry and Word Count (LIWC) with deep learning techniques for multimedia analysis, which involves leveraging the linguistic insights from text-based analysis and combining them with deep learning models capable of processing various modalities, including audio and videos. We first present some concepts and definitions related to psycholinguistic measurement. Then we talk about multimedia content analysis

Keywords: Psychometric, Multimedia Analysis, LIWC, Text Analysis, Deep learning, LLMs, Multimedia Content.

Résumé

Ce travail est consacré à l'intégration de l'investigation linguistique et du comptage de mots (LIWC) avec des techniques d'apprentissage profond pour l'analyse multimédia, ce qui implique de tirer parti des connaissances linguistiques du texte-analyse basée et intégration avec des modèles d'apprentissage profond capables de gérer divers contenus, y compris audio et vidéo. Nous présentons d'abord quelques concepts et définitions liés à la mesure psycholinguistique. Nous parlons ensuite d'analyse de contenu multimédia

Mots-clés: Psychométrique, Analyse Multimédia, LIWC, Analyse de Texte, Apprentissage Profond, LLM, Contenu Multimédia.

الملخص

هذا العمل مخصص لدمج التحقيق اللغوي وعدد الكلمات مع تقنيات التعلم العميق لتحليل الوسائط المتعددة، والذي يتضمن الاستفادة من الرؤى اللغوية من التحليل القائم على النص ودمجها مع نماذج التعلم العميق القادرة على معالجة المحتويات المختلفة، بما في ذلك المحتوى الصوتي ومقاطع الفيديو. نعرض أولاً بعض المفاهيم والتعاريف المتعلقة بالقياس النفسي اللغوي. ثم نتحدث عن تحليل محتوى الوسائط المتعددة.

الكلمات المفتاحية: القياس النفسي، تحليل الوسائط المتعددة، التحقيق اللغوي وعدد الكلمات، تحليل النصوص، التعلم العميق، نماذج اللغة الكبيرة، محتوى الوسائط المتعددة.

Table of contents

List of Tables

General Introduction

Analyzing diverse forms of multimedia content, including videos, audio recordings, and text, presents a myriad of challenges and opportunities. At its essence, multimedia analysis involves the extraction of meaningful insights from these varied forms of data, aiming to understand patterns, derive insights, and make informed decisions. The fusion of linguistic inquiry and word count (LIWC) techniques with state-of-the-art deep learning methodologies marks a significant advancement. This convergence offers a unique opportunity to delve into the intricate layers of multimedia content. By harnessing the power of LIWC's linguistic insights and coupling them with the capabilities of deep learning models, we embark on a journey to unravel the emotional, cognitive, and psychological dimensions embedded within diverse media formats.

Our problem is how to integrate Linguistic Inquiry and Word Count (LIWC) and deep learning with large language models (LLMs) within the realm of multimedia analysis, aiming to achieve a holistic understanding of multimedia content.

To address this issue, we propose an architecture that integrates multimedia, LIWC techniques, and LLMs.

This work is organized as follows:

1. The first chapter, entitled ”Psycholinguistic Measurement,” includes psycholinguistic measurement, its significance, and applications, and discusses the nature of languages.

2. The second chapter, entitled ”Multimedia Analysis,” explains the techniques, approaches, and applications of multimedia analysis, peeling light on the significance and impact of deep learning and the usage of large language models (LLMs).

3. The third chapter describes the proposed system and program design.

4. The fourth chapter presents the development environment, the programming language, results, and main interfaces that concern the application.

Chapter 1

Psycholinguistic Measurement

1.1 Introduction

The field of psycholinguistic measurement is a valuable tool in the study of human behavior, cognition, and communication. It provides psychologists and educators with information on the abilities and personalities of an individual, which in turn facilitates making proper decisions in the most diversified settings.

In this chapter, we shall look at the field of psycholinguistic measurement, its significance, and its applications. We show the use of psychological tests across various domains, including education and the clinical field, and then we discuss the tools used for psychological measurement. Furthermore, we are looking into the multilevel nature of language, focusing on Arabic and English as examples.

1.2 Understanding psycholinguistic measurement

1.2.1 Definition of measurement

In a broad sense, measurement refers to the quantitative aspects that describe a specific property or trait. Measurements that are commonly used include the height of a liquid or an individual's temperature. Accordingly, measurement is a term used to describe quantitative aspects that describe the characteristics or traits of a person, substance, or any phenomenon in a particular field. In order to understand the true concept of measurement and analyze its nature scientifically, it must be understood from different aspects, which are [1]:

- Thorndike gives his philosophical opinion on analogy in his famous saying: "If a thing exists, it exists in some amount; and if it exists in some amount, it can be measured".

- Measurement uses many tools, such as machines, verbal tests, group practical tests, and individuality, as in physical, psychological, and social tests.
- The measurement is performed either directly through devices or indirectly through tests, standards, or non-test methods.
- Quantization is how measurement is done, and it involves converting a type, attribute, or characteristic into a numerical value that is related to specific standards.
- The term quantization means converting a type into a number, instead of saying this object is long or short As a qualitative description, we describe it as a quantitative description, so we say that the length of the body is (170) or (175) or (180).
- To measure, two basic conditions must be met, one of which is a quantitative description of the object, and the second is that this amount be attributed to a specific standard.

1.2.2 Definitions of psychological measurement

Psychological measurement has multiple definitions, including [1]:

- Refers to quantitative features that classify a class of numbers or symbols that correspond to a class of qualities or events, according to well-defined rules, or characterize a specific quality or trait, like a student's academic accomplishment.
- It is the process of describing information (quantitatively), i.e., utilizing numbers to categorize, describe, and arrange data in a straightforward, objective manner that can be comprehended and subsequently analyzed.
- It's a method of quantification, which is the expression of qualities, factors, or phenomena of qualitative, moral, or behavioral issues in quantitative or mathematical terms that calls for an assessment judgment to be made about them.
- It is the procedure that allows things, characteristics, and properties to be numerically valued under specific circumstances.
- This type of assessment is quantitative, expressing measured phenomena and qualities through the language of numbers.
- It is the process of translating or transforming observed events or things into digital numbers using specialized instruments.

1.3 Characteristics of psycholinguistic measurement

Psychometric is characterized by a set of general characteristics, the most important of which are [3]:

- Psychometric measurement is a quantitative estimate of a dimension of behavior. For example, obtain grades that express the level of the student's achievement and mental abilities.
- The score that an individual obtains on a psychological test has no meaning in itself; rather, comparing it with a standard gives it meaning within which it can be understood. The basic criterion for judgment is derived from the characteristic itself. What is the meaning of saying that a child's IQ is 100 and that the score itself has no significance? But in order for it to have meaning, it must be compared to a (basic) standard of judgment derived from the nature and distribution of intelligence. Thus, the individual's level of intelligence can be determined.
- Psychological measurement is an indirect measurement, as we cannot measure intelligence, achievement, or any characteristic. Another psychological way is as direct as measuring the height or weight of individuals. Psychological measurement is similar to measuring some natural phenomena, such as temperature, as we only measure temperature by its impact on a column of mercury, that is, we measure it indirectly.
- Psychological measurement is a relative measurement, not an absolute one, as a result of the absence of a known absolute zero in physical measurement. The standards we use in psychological measurement are derived from observed behavior for a specific group of individuals under specific circumstances. This means that the interpretation of the degree to which the individual obtains it in any psychological test is only done by comparing it with the standards derived from the group to which the individual belongs.
- There are errors in psychometrics, as there are in measurement in any field of science. These errors may be due to examiners, measuring tools, or a lack of agreement about something being measured.
- Psychometrics is just a means, not an end in itself. It is only useful to the extent that it helps teachers, mentors, managers, and others improve and develop their work to the extent necessary. It helps in understanding human behavior.
- Psychometrics is the isolation of characteristics and traits. Traits do not exist in isolation from each other. In nature, they are interwoven and overlapping. Intelligence

intersects with social maturity and achievement. Academic, physical maturity, etc. Therefore, to measure intelligence, it must be isolated from other intelligence traits such that the estimates we arrive at are accurate in their quantitative expression of intelligence without other features.

- There is no single, fixed-valued, agreed-upon unit of measurement used to measure traits differently. All lengths are measured using centimeters as a unit of measurement, and all weights are measured using grams as the unit of measurement. Not all intelligence tests may use a fixed-valued unit. The most famous unit may be the number of points obtained by the examinee according to certain rules. Lack of agreement on a specific unit of measurement increases the relativity of psychological measurement; on the other hand, it does not help to compare the performance of one individual on two different tests. A child's intelligence, as measured by the Wechsler test, is not the same as that measured by the Cattell intelligence test.

- Psychological measurement is less accurate than measuring natural phenomena. This means that if we measure a person's intelligence, then we measure his intelligence again two weeks later, and we do not get the same score, which was close to first class. However, this does not happen when measuring natural phenomena. Under normal conditions, water freezes above zero or below 100 degrees, no matter what we repeat in the measurement process. This is because the psychological phenomenon is affected by many factors that may not be present. It can be controlled, so it is impossible to obtain the same estimate when measuring more than once. Then Measuring a person's intelligence for the second time is based on his familiarity with the test, his motivation, and his relationship with the examiner. His health conditions and other factors affect his performance on the test. So he strives to ensure that the signers of measurement methods require that their instruments have certain characteristics so that maximum performance can be achieved with a degree of accuracy in their tools (conditions for good testing).

From the above, it is clear that when we are in a situation in which a decision is required for an individual to determine his future or determine his fate, psychological tests have a very important role, and they should never be the only or decisive role. We must remember that there are societal standards that differ from each other. From one society to another, there are limits to the reliability of the measures, and the validity of the measures is not decisive and conclusive in all areas. All of these things make us not use decisions that are based on the results of one measure or even several measures. A decisive decision that is not subject to doubt or falsehood.

1.4 Areas of psycholinguistic measurement

Psychological and educational tests are applied in many fields to analyze an individual's abilities and talents, aptitudes and inclinations, and learn about many different aspects of the individual's personality. that The practicing psychologist must understand what tests are, what they entail, and the topics and limitations of their use The framework of the clinical and educational method and, in light of his understanding of the general principles that must be prepared and applied, psychological testing. A wide variety of tests and scales are available for psychology To choose from according to the requirements of the situation. The specific test must be gender-appropriate The subject, his age, and his mental, cultural, educational, and social level... so that these do not differ Characteristics about the characteristics of the group on which the test was based. Among the areas of psychological measurement are the following [3]:

1.4.1 Psychometric in the educational field

In the field of education, tests are utilized to provide educational guidance by measuring students' abilities. Based on their academic inclinations and aptitudes, educational administration can then distribute students into appropriate educational programs that suit their abilities, aptitudes, inclinations, and general intelligence. This process ensures that students are placed in studies that they are inclined towards, enabling them to succeed and make progress, leading to their good adaptation and comfort.

The importance of educational and psychological measurement in school life cannot be understated, as it helps prevent feelings of failure and frustration. In addition to its role in educational administration, it is responsible for categorizing students into different educational programs that match their abilities, aptitudes, inclinations, and general intelligence. Teachers utilize various psychological and educational tests and standards to group students homogeneously based on their intelligence or special abilities, allowing them to apply appropriate teaching methods tailored to each group's level.

Educational and psychological measurement also plays a crucial role in confirming the evaluation of students' work and achievements, as well as assessing the impact of different teaching methods employed by teachers. Teachers may investigate factors influencing the achievement process, such as intelligence, psychological adaptation, emotional balance, family circumstances, or health conditions. By conducting psychological tests, teachers can uncover the relationships between these factors and student achievement.

1.4.2 Psychometric in the professional field

Various psychological tests and measures are applied in vocational guidance, vocational selection, Training and vocational qualification.

- **Vocational guidance:** this means directing the individual to a type of profession in which he is likely to achieve the greatest amount of success and progress. This means that in career guidance, we choose for the individual, from a large number of professions, one profession, so that this profession is more compatible with his abilities, aptitudes, inclinations, and intelligence. This professional guidance is based on studying the individual's personality using many methods, including personal tests and interviews so that you get a true and comprehensive picture of the individual's personality.

- **Vocational choice:** in vocational choice, we are faced with a large number of individuals applying for a specific job, and we choose it from among the individuals who suit it.

- **Vocational training:** it is a type of education or acquisition of skills, experiences, and knowledge. Measurement is used to identify people for a specific type of training, that is, to predict their success and benefit from the training it provides them.

- **Vocational rehabilitation:** this means training individuals or people with disabilities and infirmities in the work that they do. It is proportional to their abilities, talents, and aptitudes, and this means that it is a type of training or education, but it also restores the individual's psychological adaptation in addition to his professional readjustment.

1.4.3 Psychometric in the clinical field

Psychological tests are used in clinical settings, i.e., in hospitals and psychiatric clinics. To know the type of psychological disorders and diseases that the patient suffers from, based on the application Tests can diagnose the disorder, and then treatment plans and programs can be drawn up. And it is not limited. Psychological measurement in the field of treatment focuses on the diagnosis, but it also includes knowing the patient's abilities and intelligence to determine the extent of the impact of these factors on his disorder and the extent of their use in his adaptation to life. In psychological diagnosis and in interpreting patient behavior, statistical methods are derived from objective tests, such as the "Multifaceted Personality" test, in addition to projective ones, such as "ink blots" and "I understand the topic," depending on the psychologist's experience. To measure In addition to the usefulness of tests in clinical fields, these tests are also used to assess mental weakness or to classify individuals into homogeneous groups. Classification is one of the important goals that can be achieved by applying tests. This classification is used when students are divided into homogeneous groups. In terms of how much intelligence and abilities they have.

1.5 Psycholinguistic measurement tools

1.5.1 Linguistic Inquiry and Word Count: LIWC2015

The latest development of LIWC2015 has brought about substantial changes to both the word list and the software choices. Significantly, the LIWC2015 software and dictionary represent a fresh approach, not simply a minor revision of past LIWC versions. Like with past iterations, the software is created to analyze and effectively examine single or multiple language files. Simultaneously, the program strives to maintain transparency and adaptability in how it functions, enabling the user to investigate word usage through various methods.

The heart of the text analysis strategy is the LIWC2015 dictionary, which consists of close to 6,400 words, word stems, and specific emoticons [28].

1.5.2 Linguistic Inquiry and Word Count: LIWC-22

The fundamental components of LIWC-22 are software and a ”dictionary”, which is a map that links significant psychological theories and constructs with words, sentences, and other linguistic creations. LIWC-22 is designed to accept written or transcribed spoken text that has been stored digitally. The topic modeling function of LIWC-22 was developed with mindfulness of a psychological method known as the Meaning Extraction Method (MEM). The MEM enables users to apply factor analyses to their text analyses to uncover primary themes and meanings within their dataset.

The LIWC-22 dictionary is the core of the text analysis approach. The internal dictionary has more than 12,000 words. In the most recent edition of LIWC, a few new categories have been created, some have undergone significant overhauls, and a small number have been eliminated [7].

1.6 Language concept

The primary way to communicate is through language, which encompasses thought and behavior, not just verbal symbols. Language is a fundamental part of culture and the humanities and is closely associated with thought. Indeed, the essence of any spoken word is a thought. Language has the power to develop or disappear, and it is the way a person

expresses their identity and culture. Bin Jinni defined language as representing the cultural identity of any society; they are not just sounds but rather an expression of the intellectual data intended by these symbols.
The French writer Ferdinand de Saussure is considered one of the first Western writers to explore language after Ben Jinni. Saussure defined language as a social phenomenon that is a set of signs and symbols that individuals use to express their entire thoughts[2].

1.7 Arabic language

1.7.1 Definition of Arabic language

Arabic is a language of the Semitic family. They include many Arabic dialects: Ugaritic, Hebrew, Phoenician, Aramaic, Syriac, Ethiopic, South Arabic, and Ak-Kadian (Babylonian and Assyrian). A contiguous region encompassing Ethiopia, the Arabian Peninsula, and the Fertile Crescent has been the permanent home of Semitic languages since the beginning of recorded history. In this domain, researchers have delineated three principal geographic dispersions of Semitic languages. The earliest Semitic language known to history, Akkadian, is a language of Northeast Semitic origin, spoken in Mesopotamia and dating to the third millennium B.C. Although the dialect was eventually supplanted by Babylonian and Assyrian, Akkadian got its name from the capital city of Akkad. Northwest Semitic languages originated in the Syria-Palestine region, where languages such as Canaanite, Phoenician, Hebrew, Moabite, Aramaic, Syriac, Nabataean, Palmyrene, and Mandaean were spoken. The earliest known inscriptions from the Northwestern region date to the middle of the second millennium B.C. The Southwest Semitic group includes the regions of Arabia and Ethiopia, and its two principal languages are Arabic and Ethiopic. There are dialects of Arabic that are spoken in the north and south. The first inscriptions in Southern Arabic may be traced back to the eighth century B.C. and include knowledge of the kingdoms of the Sabaeans, Minaeans, Qabatanians, and Himyarites. Northern Arabic appeared much later, although Thamudic, Lihyānite, and Ṣafaitic inscriptions may be related to early Northern Arabic. We did not know anything about poetic Arabic until the sixth century A.D., although it seems to have evolved into the language of the Quran in the century that followed. This Arabic is the Arabic of Islamic times, or what is commonly called ”Arabiyyah.”

The Arabic script has twenty-eight letters standing for consonants and is written from right to left. Many of these are identical except for the dots above or below the letter. Each letter has a somewhat different shape depending on whether it occurs at the beginning, middle, or end of a word. Some of the consonant sounds do not have equivalents in Western languages. The Arabs like to refer to themselves as al-nāṭiqun bi-l-Dad, literally ”the speakers with the letter D,” because an emphatic d is believed to be a unique feature of the language. In addition, the alphabet has three symbols for short vowels: dammah (u), fathah (a), and kasrah (i). These and other signs are written as diacritical marks. There are also three corresponding long vowels, ā, ā, and i, which are expressed with individual characters and written as integral parts of a word. The short vowels have never been made a permanent part of the writing system, with the single exception, perhaps, of the Qur'an, in which the vowels are always written to ensure correct reading of the Divine Revelation. It should be added that in learning to read the language, vocalized texts are used for some time, after which one must learn to read without vowels. Such an orthography makes for faulty reading, even on the part of those who possess a thorough knowledge of Arabic morphology (sarf) and grammar (nahw). This difficulty is generally recognized, although not candidly admitted, by those who profess to know the language. Mistakes in reading inevitably occur unless the reader understands beforehand what he is about to read. Moreover, since literary Arabic is an inflected language with endings, which are rarely written, for the nominative, accusative, and general cases of nouns, reading is rendered all the more taxing [10].

1.7.2 Fundamentals of Arabic language

More than a dozen states speak modern Arabic, which is spoken by 80 million people. Arabic has a long history of being one of the great languages of civilization. It traveled with Islam through North Africa, the coast of East Africa, up into Central Asia, and throughout Southeast Asia as a liturgical language. During the Middle Ages, Arabic helped to preserve Greek science and developed a rich literary tradition that is most known in Europe from the Arabian Nights. Formally speaking, Arabic is the most significant language of the Semitic group, with a complex and unusual structure, and it has no relation to English at all.

The Arabic language, which is the primary written form today and has served as the vehicle for Islam and literature, is relatively uniform throughout the Arab world and the Islamic centuries, but it has never been the spoken language of the Arab people. This unique social context surrounding the use of Arabic must be understood to address the subject. The language of everyday communication is colloquial Arabic, which reflects the diverse linguistic, socioeconomic, and religious backgrounds of the populace. Diglossia is the term for this condition, which causes issues for description as well as for the Arabs themselves [6].

1.8 English language

1.8.1 Definition of English language

English finds its origins in the Germanic dialects spoken by tribes in northwest Europe, including regions of contemporary northwest Germany, during a subsequent migration wave where numerous modern Germanic languages contain words shared with English. After the Romans vacated Britain in AD 410, the Anglo-Saxons made their entry into the country in AD 449, as documented in conventional historical narratives. Throughout Britain's history, a diverse array of peoples, encompassing the Britons, Romans, Anglo-Saxons, Scandinavians, and French, have graced its shores, each bringing with them their distinct languages. This rich linguistic mosaic has exerted a profound influence on the development of the English language. Delving into the etymology of place names affords us a deeper understanding of the multifaceted linguistic influences that have left their mark on English across different epochs [13].

More than 300 million people worldwide speak English as their first language, and many more use it as their second language. In trade, diplomacy, sport, science, technology, and numerous other fields, it is used to communicate internationally [32].

1.8.2 Structure of English language

Grammar in English

The study of grammar focuses on the fundamental ideas of language. English grammar teaches us how to speak and write the language correctly by addressing its tenets and usages. It is related to [19]:

- the language's fundamental letters and sounds
- the way words are categorized and changed
- the way sentences are put together
- the rules governing its versification.

Grammar is therefore separated into four categories: etymology, syntax, and prosody.

- Orthography deals with basic sounds, the letters that stand for them, and how to combine letters to form words and syllables. The correct pronunciation of words is treated in orthopedics.
- Etymology studies the categorization, origin, and diverse changes made to words.

- Sentence structure is the subject of syntax.
- Poetry discusses the rules of versification.

Parts of speech

All language words are categorized into classes known as parts of speech. Every speech component has unique qualities. The differences between the components of speech are as follows [40]:

- meaning
- form
- function.

The English language's vocabulary can be categorized into three primary groups:

1. **Notional words:** serve as principal or secondary components of the phrase, each with a unique lexical meaning and autonomous syntactic roles.
   (a) VERB: go, sit, play...
   (b) NOUN: man, girl, table...
   (c) ADJECTIVE: nice, brilliant, clever
   (d) NUMERAL: one, twenty, fifth, tenth...
   (e) PRONOUN: I, you, my, his, somebody, this
   (f) ADVERB: nicely, often, seldom, here, there
   (g) WORDS OF CATEGORY OF STATE: asleep, alone, alive
2. **Functional words:** semantically distinct from notional words, their lexical meaning is broader than the notional words'. While structural words don't carry out any independent grammatical function in the phrase, they do convey different relationships between the words in the sentence or provide context for words like ”a” and ”the” books.
   (a) ARTICLE: a, the
   (b) PARTICLE: to, too, not, also...
   (c) PREPOSITION: in, on, with, of...
   (d) CONJUNCTION: and, but, if, though, or...

3. **Independent elements:** are terms that possess a peculiar meaning of several kinds. They are not related grammatically to the sentence in which they appear. They carry out no syntactic duties. Example:
   (a) INTERJECTION: Alas! Wow! Ouch! Oh!
   (b) MODAL WORDS: indeed, probably, no doubt
   (c) WORDS OF AFFIRMATION AND NEGATION: yes, no

Types of sentences

A sentence can include one or more assertions, and each proposition can either be independent or reliant on one or more other propositions. Three different categories of sentences result from this combination of freedom, which are categorized based on the quantity and nature of the propositions they contain. They are compound, complicated, and simple [24].

1. A sentence with only one dependent proposition is considered simple. ”Lightness of touch is the crowning test of power.” Higginson.

2. A statement is considered complex if it has at least one dependent proposition. It may have any number of dependent propositions in addition to the complete independent statement that it typically comprises.

   - With the independent proposition full, ”What inspiration gilds his features as he descends the mount with the tables in his hand.” Lord.
   - Since the dependent proposition is the topic of the independent proposition, it is incomplete, ”That Chaucer, being in Milan, should not have found occasion to ride across so far as Padua for the sake of seeing the most famous literary man of the day is incredible.” Lowell.

3. A sentence is considered compound when it contains two or more separate propositions or one or more dependent propositions, in which case it is often called a complex compound.

   - A compound phrase ”It is a strange tale, but it hath the recommendation of brevity.” Jerold.
   - Compound complex phrase ”Times of heroism are generally times of terror, but the day never shines in which this element may not work.” Emerson.

1.9 Conclusion

In conclusion, we find that psycholinguistic measurement serves as a cornerstone in understanding human behavior and cognition. Through the application of psychological tests, educators, clinicians, and researchers gain invaluable insights into individuals' abilities, aptitudes, and psychological well-being.

Chapter 2

Multimedia Analysis

2.1 Introduction

Multimedia data is becoming commonplace in the digital age, with text, audio, video, and image files serving as the foundation for communication and information sharing. Because of this, multimedia data analysis has become very important in a variety of fields, including healthcare, security, education, and entertainment. In this chapter, we explore multimedia data and the techniques and uses of multimedia analysis, highlighting the importance and effects of deep learning as well as the applications of Large Language Models (LLMs).

2.2 Multimedia

2.2.1 Definition

Multimedia combines two or more media on a personal computer, such as text, graphics, photos, audio, animation, or video. Multimedia is distinct in a few respects. First, the digital realm of the computer brings together traditional media kinds, which have historically had disparate means of storage, transportation, and presentation. Second, multimedia is interactive; by hitting a key, using the mouse, or pointing at the screen, the user can manage how information is accessible and shared. Finally, genuine media integration is provided by multimedia. Not only are many media formats shown concurrently and with equal ease, but they are also linked together so that access to them need not be done in a straight path [22].

2.2.2 Types of multimedia data

Multimedia provides a range of media typically integrated in a meaningful way. This allows for the utilization of the computer for showcasing ideas through various methods, such as [5]:

- **Images:** which encompass digitized photos, illustrations, charts, and slides.
- **Videos:** including detailed processes and interviews.
- **Audios:** like voice recordings, sound effects, and music.
- **Textual Data:** which refers to any information presented in written form.

2.3 Multimedia analysis

2.3.1 Definition

Multimedia analysis is the process of analyzing various types of multimedia data, such as images, videos, audio recordings, and other forms of multimedia content. The process of this analysis usually involves extracting meaningful information, identifying patterns, and understanding the data to gain insights or make decisions. Multimedia processing often involves live spatiotemporal data that has the following attributes [36, 11]:

- They possess a tremendous amount of redundancy.
- The data is dynamic and has temporal variations.
- It does not exist in isolation it exists in its context with other data. For instance, visual data comes along with audio, music, text, etc.

2.3.2 Techniques and approaches in multimedia analysis

Image analysis

Image analysis is used to extract information contained in images. Image analysis aims to detect, diagnose, and describe the texture and geometry of images. The processes used to generate and acquire an image determine its information content and meaning. features through which an image is identified and categorized are of spectral and geometric (or morphometric) nature [34].

Image analysis encompasses a wide range of areas across various fields. Some prominent areas include:

- **Object detection and recognition:** object detection is a long-term and challenging task in computer vision. As one of the primary tasks of image analysis, the ultimate goal of object detection is to give the classes and locations of objects within images, including faces, vehicles, animals, and everyday objects. Therefore, object detection is the basis for many other computer vision and image analysis tasks, such as object tracking, scene understanding, surveillance, autonomous vehicles, and augmented reality [26].

- **Semantic segmentation:** semantic segmentation assigns a semantic label or category label to each pixel of an image, many real-world applications benefit from this task, such as self-driving vehicles, pedestrian detection, therapy planning, medical image analysis (e.g., identifying organs or tissues), and computer-aided diagnosis [20].

- **Image classification:** image classification is a paramount topic in artificial vision systems, which aims to classify an image, which is an input, into predefined classes or labels based on its visual content. This is used in applications such as medical imaging (e.g., classifying tumors), and satellite imagery [31, 4, 8].

Video analysis

Video analysis is defined as any method or/and technique that can extract high-level information from video data, and it requires an abstraction process that is not explicitly stated [9].

There are various areas of video analysis such as:

- **Surveillance and security:** intelligent surveillance is an important use of video analysis and comprehension. Its goal is to automatically evaluate human behavior and identify anomalous events that can endanger public safety and security [18].

- **Sports video analysis:** video processing has found many applications in sports such as [35]:
  - Tactics analysis: one of the major aims of sports video analysis is to assist in training. There is a need to summarize the play tactics from video streams.
  - Tracking: is one of the most frequently used techniques in Sports analysis. It has been used in tracking balls, players, Referees, etc.
  - Highlight extraction: his goal is to give viewers a brief overview of the game without requiring them to watch an extended video broadcast.

- **Education and learning:** creating interactive educational content, simulators using video analysis techniques. This could involve analyzing user interactions, providing feedback, and adapting content based on learner behavior [12].

Audio analysis

Audio analysis encompasses any method capable of obtaining high-level information from audio signals, which is information that does not explicitly state and requires an abstraction process. The wide range of possible audio sources and the multifaceted nature of audio signals result in a variety of distinct audio problems, which lead to various areas of research, including [25, 17, 9]:

- **Speech analysis:** covering topics such as automatic speech recognition or recognizing emotion in speech.
- **Urban sound analysis:** is used for monitoring noise pollution and audio surveillance, which involves the detection of dangerous events.
- **Industrial sound analysis:** such as monitoring the state of mechanical devices like engines or monitoring the health of livestock, and, last but not least.
- **Musical audio analysis:** aims to understand and extract musical parameters and properties of the audio signal.
- **Audio event detection:** involves detecting audio events in audio streams. There can be numerous related applications, such as audio-based surveillance, violence detection, and intrusion detection, to name but a few.
- **Speaker identification, verification, and derivation:** these speaker-related tasks focus on designing methods that discriminate between different speakers. Speaker identification and verification can be useful in the development of secure systems and speaker derivation, as they can answer the question 'who spoke when? ', can be used in conversation summarization systems.

Text analysis

Text analysis refers to any method or technique that can extract high-level information from texts, and it is a process that is not explicitly stated. Text information is often available alongside the image/video/audio content; features can be extracted from the transcripts obtained with automatic speech recognition (ASR) or closed caption (CC), optical character recognition (OCR), and production metadata. Techniques for extracting such features are similar to those in text retrieval, such as word counts in a bag-of-words representation. Besides text-only features, speech signals have timing information that can determine the speaking rate and pause length [14].

There are various areas for text analysis, including:

- **Sentiment analysis:** determining the sentiment or opinion expressed in text data, whether it's positive, negative, or neutral. This is widely used in social media monitoring, customer feedback analysis, and market research [33].
- **Text classification:** is a construction problem of models that can classify new documents into pre-defined classes. Applications include spam detection, topic categorization, and news article classification [27].
- **Automatic text summarization (ATS):** produce a summary that includes the main ideas in the input document in less space. This is useful for news aggregation, document summarization, and automatic report generation [15].

2.4 Deep Learning in multimedia analysis

2.4.1 Definition

The branch of artificial intelligence known as ”deep learning” is dedicated to building huge neural network models that can make precise judgments based on data. Deep learning works best in situations where there are lots of available datasets and complex data. Deep learning is used in the majority of online businesses and high-end consumer technology today. Additionally, Facebook analyzes text in online interactions using deep learning. Deep learning is used by Microsoft, Baidu, and Google for both machine translation and image search. Deep learning algorithms are now standard on all new smartphones; for instance, deep learning is used for face detection on digital cameras as well as speech recognition. Deep learning is used in the healthcare industry to process medical pictures (such as X-rays, CT, and MRI scans) and make diagnoses. Deep learning is also the brain behind self-driving cars, where it powers motion planning and steering, environment perception, localization and mapping, and driver status tracking [23].

2.4.2 Deep Learning architectures for multimedia analysis

Convolutional Neural Network (CNN)

A CNN is a kind of neural network used in AI, NLP, speech, and images. This network was first introduced by Hubel and Wiesel. These networks are considered deep neural networks consisting of three hidden input and output layers, the convolution neural network center has a convolution layer, which is why this type of network is named. The input layer is an array of numbers. Image Net database data developed this architecture using the Relu activator function. Due to the high performance speed and reduced training time, these networks have high accuracy in image recognition, And feature recognition without human intervention. The high number of input data can be a disadvantage of this network [30].

Recurrent Neural Network (RNN)

This type of network, also called RNN, was first created by David Rumelhart in 1986 and developed later. It is a kind of neural network used in speech recognition NLP and sequential data processing. These networks do not transmit information in just one direction (from the input layer to the output one). In RNN, each node acts as a memory cell and carries on operating and calculating. Despite Feed Forward Neural Networks, in RNNs, edges can form circles. These networks can easily remember their previous input because of their memory power and use it to process the following sequence. LSTM is one of the most popular types of these networks. RNN is used for time series, text, and audio data. This network cannot memorize the sentence for long-term, that is the weakness of the networks, which is the so-called Vanishing gradient problem. This network can be combined with the convolutional networks, The problem with this network is that it is difficult to train. The performance of RNN is such that it is necessary to know the previous words to predict the next term. RNN repeats the operations as is evident from its name (implies the same operation on all parts of a sequence (or series) of inputs) [30].

Transformers

Transformers are a form of deep neural networks that address the drawbacks of sequence-to-sequence (seq-2-seq) structures, such as the restricted sequence input dependency and the linear input processing that impedes simultaneous training of networks. Transformers utilize the multi-head self-attention method for extracting features, showing significant promise for use in NLP. Instead of traditional methods of repetition, Transformers use attention to understand an entire section of a sequence by utilizing encoding and decoding blocks. Transformers have a significant benefit compared to LSTM and RNNs due to their attention mechanism, allowing them to grasp the genuine context meaning. Additionally, Transformers have speed advantages as they can operate simultaneously, unlike recurrent networks, and can be processed using Graphical Processing Units (GPUs), enabling quicker completion of tasks with extensive inputs. The benefits of the Transformer model have motivated deep learning experts to investigate its possibilities for diverse tasks in a variety of fields, resulting in many research papers and the creation of Transformer-based models for multiple tasks in the realm of artificial intelligence [21].

2.5 Large Language Models (LLMs)

2.5.1 Definition

A large language model is a language model with massive parameters that undergoes pretraining tasks (e.g., masked language modeling and autoregressive prediction) to understand and process human language, by modeling the contextualized text semantics and probabilities from large amounts of text data. A capable LLM should have four key features:

1. Profound comprehension of natural language context.
2. Ability to generate human-like text.
3. Contextual awareness, especially in knowledge-intensive domains.
4. Strong instruction-following ability which is useful for problem-solving and decision-making.

Large Language Models (LLMs) represent an evolution from language models. Initially, language models were statistical and laid the groundwork for computational linguistics. The advent of transformers has significantly increased their scale. This expansion, along with the use of extensive training corpora and advanced pre-training techniques is pivotal in areas such as AI for science, logical reasoning, and embodied AI. These models undergo extensive training on vast datasets to comprehend and produce text that closely mimics human language. Typically, LLMs are endowed with hundreds of billions, or even more, parameters, honed through the processing of massive textual data. They have spearheaded substantial advancements in the realm of Natural Language Processing (NLP) and find applications in a multitude of fields (e.g., risk assessment, programming, vulnerability detection, medical text analysis, and search engine optimization).

Large Language Models (LLMs) like ChatGPT and Bard have revolutionized the understanding and generation of natural language. They possess deep language comprehension, human-like text generation capabilities, contextual awareness, and robust problem-solving skills, making them invaluable in various domains such as search engines, customer support, and translation [37].

2.6 Popular LLMs

2.6.1 ChatGPT

ChatGPT, developed by OpenAI, is famous for its capacity to hold conversations and generate creative text. Developed using the Generative Pre-trained Transformer (GPT) architecture, this technology predicts the following word in a given text, allowing for smooth and coherent text creation.

ChatGPT utilizes a combination of Reinforcement Learning with Human Feedback (RLHF) to enhance responses using human input and employs instruction tuning for flexibility. ChatGPT, which emphasizes natural language interactions, benefits from the reinforcement learning-based technique of reducing bias. This leads to responses that are safer and more useful. Nevertheless, OpenAI's level of transparency about training data has been called into question in contrast to Google's more thorough approach [29].

2.6.2 Gemini

Gemini, developed by Google AI, comprises a family of LLMs, with Gemini Ultra 1.0 being the most sophisticated version. One important aspect is the integration of information retrieval and text generation, known as Retrieval-Augmented Generation (RAG), which ensures that the generated outputs are rooted in factual information.

Gemini is trained on a wide range of datasets, enabling it to perform well in different tasks and take advantage of Google's extensive infrastructure. Gemini is very good at providing accurate information through RAG, but there are times when it may create harmful content from its database. It excels in performing a wide range of tasks, including code generation, outperforming ChatGPT in terms of task variables [29].

2.6.3 Architectural aspects of Gemini and ChatGPT

Gemini and ChatGPT, two prominent large language models (LLMs), have become widely acknowledged for their impressive abilities in text generation. Nevertheless, their architectural approaches diverge significantly, impacting their respective strengths, weaknesses, and potential outcomes for LLM development. Table 2.1 displays the architectural features of Gemini and ChatGPT [29].

Table 2.1: Architectural aspects of Gemini and ChatGPT

Feature	Gemini	ChatGPT
Developer	Google AI	OpenAI
Model Type	Multimodal Language Model (can handle text, images, and potentially other modalities)	Generative Pre-trained Transformer (GPT) Language Model
Base Architecture	Believed to be a Transformer-based architecture	Transformer-based architecture

Table 2.1 Architectural aspects of Gemini and ChatGPT

Feature	Gemini	ChatGPT
Model Variations	Available in three sizes, the model comes in Ultra, Pro, and Nano variants. Ultra represents Google's most capable model, while Pro offers versatility for a wide range of applications. Nano, on the other hand, is a more focused and compact model designed for specific tasks.	Starting with GPT-1, the series has evolved into through several iterations, including GPT-2 and GPT-3, leading up to the more recent GPT-4, which powers ChatGPT and continues to push the boundaries of AI language models.
Training Data	Extensive, proprietary dataset meticulously curated by Google, possibly incorporating diverse data types such as text, code, and more.	A colossal repository comprising both text and code-based data, meticulously filtered to enhance conversational quality and ensure safety.
Key Strengths	Advanced comprehension of language with the added advantage of multimodal capabilities, expanding its potential for diverse applications. The range of model variations enables customization to meet specific requirements.	Outstanding conversational skills paired with the ability to generate diverse and imaginative text formats. Proficient in adhering to instructions and maintaining relevance to the topic at hand.
Known Weaknesses	Further details are somewhat scarce due to limited accessibility. There is a possibility of similar biases and the potential to generate misinformation, akin to other large language models.	Encounters challenges in certain domains of logic and reasoning. Additionally, it can be manipulated with strategic prompts to generate undesired outputs.

2.6.4 Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis

ChatGPT and Gemini stand out as top-tier large language models (LLMs) within the field of artificial intelligence. Renowned for their prowess in natural language communication, imaginative text generation, and insightful responsiveness, these models have garnered significant acclaim. ChatGPT, developed by OpenAI, has emerged as a widely recognized entity, while Gemini, a creation of Google, presents a formidable competitor with its distinctive attributes. A thorough examination employing the strengths, weaknesses, opportunities, and threats (SWOT) framework provides valuable insights into the unique advantages, constraints, and prospective trajectories of both ChatGPT and Gemini in the realm of AI. Below in Table 2.2, a comprehensive delineation of the SWOT analysis for both models [29].

Table 2.2: SWOT analysis of Gemini and ChatGPT

SWOT	**Gemini (Google)**	**ChatGPT (OpenAI)**
Strengths	Handles multiple modalities such as text and images: Capable of processing and generating responses using a variety of input formats.	Continues to undergo development: May encounter irregularities and constraints when compared to more established models.
	Enhanced reasoning logical abilities: Demonstrates improved comprehension and logical reasoning in its output.	Limited real-world data compared to ChatGPT: Possibly less robust factual knowledge and nuanced understanding.
	Computational efficiency: Operates effectively with minimal computational resources, potentially enhancing speed and accessibility.	Potential misuse of enhanced capabilities: Open access heightens the risk of exploitation for deceptive or malicious intents.
	Open-source commitment: Offers transparency and the possibility for community-driven enhancements and ethical oversight.	Resource intensive: Requires significant computational resources, which could pose challenges for widespread deployment.
Weaknesses	Limited multimodal capabilities: Mainly oriented towards text, may face challenges in matching Gemini's proficiency across different data formats.	Facing competition from other emerging LLMs: Pressure to stay at the forefront of the rapidly advancing AI landscape.
	Risk of bias and dissemination of misinformation: Could propagate incorrect or harmful information due to its training data.	Negative publicity surrounding AI misuse: Instances of high-profile misuse may erode public confidence in the capabilities and ethical usage of AI technologies.

Table 2.2 SWOT analysis of Gemini and ChatGPT

SWOT	Gemini (Google)	ChatGPT (OpenAI)
	Potential biases within datasets: Unchecked biases in training data have the potential to permeate AI system responses, leading to unfair outcomes and reinforcing existing societal prejudices.	Resource intensive: Requires significant computational resources, which could pose challenges for widespread deployment.
Opportunities	Revolutionize multimodal user experiences: Enable interfaces to integrate text-based, voice-based, and visual interactions.	Ethical considerations of multimodal AI: Accountability for the unintended outcomes arising from the amalgamation of diverse media formats, such as the generation of potentially detrimental content.
	Tackle intricate problems through logical analysis: Assist by employing logical deduction and complex problem-solving techniques.	Regulation of AI technologies: Stringent regulations and standards might hinder the progress of AI innovation, stifling the advancement of technology and limiting its potential benefits.
	Promote innovation through an open-source model: Encourage collaborative model development, and customization, and emphasize ethical deployment.	Potential for commercialization: Provides avenues for licensing agreements and specialized service offerings.
Threats	Facing competition from other emerging LLMs: Pressure to stay at the forefront of the rapidly advancing AI landscape.	Negative publicity surrounding AI misuse: Instances of high-profile misuse may erode public confidence in the capabilities and ethical usage of AI technologies.
	Ethical considerations of multimodal AI: Accountability for the unintended outcomes arising from the amalgamation of diverse media formats, such as the generation of potentially detrimental content.	Public dependence on AI systems: Potential risks associated with AI systems encompass algorithmic errors, susceptible to opinion manipulation, and the peril of excessive reliance on generated outputs.

Table 2.2 SWOT analysis of Gemini and ChatGPT

SWOT	Gemini (Google)	ChatGPT (OpenAI)
	Potential biases within datasets: Unchecked biases in training data have the potential to permeate AI system responses, leading to unfair outcomes and reinforcing existing societal prejudices.	Regulation of AI technologies: Stringent regulations and standards might hinder the progress of AI innovation, stifling the advancement of technology and limiting its potential benefits.

2.6.5 Comparison of popular LLMs

As shown in Table 2.3, there is a diversity of providers for language models, including well-known companies like Google, OpenAI, and Meta AI as well as up-and-coming companies like Anthropic and Cohere. The releases, which take place between 2018 and 2023, demonstrate how quickly language models have advanced and changed in recent years. In 2023, newer models like ”gpt-4” have surfaced, demonstrating the continued progress in this industry. It’s interesting to note that some models, such as BERT, T5, PaLM, LLaMA, and CTRL, are open-source, despite the fact that the majority of the models are not. This can support community-driven research and applications. More parameters are typically found in larger models, which may indicate more capabilities but also higher computing needs. For instance, ”PaLM” sticks out due to its enormous 540 billion parameters. Additionally, it is noted that LLMs typically have a higher number of parameters, which may indicate stronger capabilities but also higher computing demands. The column labeled ”Tunability” indicates if these models are amenable to being adjusted for particular tasks. Put differently, a big, pre-trained language model can be trained on a smaller, domain-specific dataset, and its parameters can be adjusted to improve its performance on a given task. For example, BERT’s tunability allows for the fine-tuning of the algorithm to achieve exceptional sentiment analysis on a dataset of movie reviews [37].

Model	Date	Provider	Open-Source	Params	Tunability
gpt-4	2023.03	OpenAI	X	1.7 T	X
gpt-3.5-turbo	2021.09	OpenAI	X	175 B	X
gpt-3	2020.06	OpenAI	X	175 B	X
cohere-medium	2022.07	Cohere	X	6 B	✓
cohere-large	2022.07	Cohere	X	13 B	✓
cohere-xlarge	2022.06	Cohere	X	52 B	✓
BERT	2018.08	Google	✓	340M	✓
T5	2019	Google	✓	11 B	✓
PaLM	2022.04	Google	✓	540 B	✓
UaMA	2023.02	Meta AI	✓	65 B	✓
CTRL	2019	Salesforce	✓	1.6 B	✓
Dolly 2.0	2023.04	Databricks	✓	12 B	✓

Table 2.3: Comparison of popular LLMs

2.6.6 Applications of LLMs in multimedia analysis

Image Captioning

Image Captioning generates descriptive sentences from images using Vision-Language Pre-trained models (VLPs). This task can be approached using LLMs by understanding the content of the image and expressing it in natural language [38].

Video captioning

Video Captioning is a task that involves generating natural language descriptions of the video content. This task can be approached using LLMs by training them on a large dataset of videos with corresponding captions. Video captioning using LLMs finds application in various areas, including enhancing accessibility for individuals with hearing impairments, facilitating video search and retrieval, generating video summaries, and improving overall understanding of video content [39].

Video retrieval

Video Retrieval using LLMs refers to the process of searching and retrieving relevant videos from a large video database using advanced language models. This task can be approached using LLMs by training them on a large dataset of videos with corresponding textual descriptions [39].

Video question answering

Video Question Answering is a task that involves answering natural language questions about the video content. This task can be approached using LLMs by training them on a large dataset of videos with corresponding questions and answers. The model learns to extract relevant information from the video content to answer the question. The advantage of this approach is that it can generate specific answers to specific questions [39].

Content moderation

LLMs assist in moderating multimedia content by analyzing the content (e.g., as it is uploaded) or reports of users who visualize the content on the platforms (e.g., private conversations, etc.) or experience potentially dangerous situations (e.g., online grooming, spam, blackmailing, etc.). They identify inappropriate or harmful content, keeping platforms and their users safe from malicious activities and harmful content [16].

2.7 Conclusion

In conclusion, multimedia analysis takes the forefront in decision-making and insight generation in digital world today. We can discover the power of multimedia data through video analysis, audio analysis, and text analysis using state-of-the-art techniques and approaches to obtain valuable information and to further developments in different fields. Furthermore, deep learning and the development of large language models are changed the scope of analyzing multimedia entirely, whereby processing and understanding of the information contained in the multimedia is possible. In the development of the power of analyzing multimedia, we open wider communication, better decision-making, and a better user experience in the digital domain.

General Conclusion

Multimedia analysis is an important field in today's digital age in many areas, as it facilitates the understanding of content and the extraction of important information. This field constitutes the processing of text, audio, video, and images to extract meaningful insights that shall lead to better decision-making and support functionalities such as content recommendation, security monitoring, and automated transcription. With the help of leading algorithms and techniques in deep learning, the analysis of multimedia contributes to a great extent to the new media, communication, entertainment, healthcare, and various other sectors, which eventually allow a deep understanding of complex data.

In our work, we integrate Linguistic Inquiry and Word Count (LIWC), deep learning, and Large Language Models (LLMs) to analyze different multimedia contents, and that's what makes it the first work in the field of LLMs.

Our work is presented as follows:

- Extract textual data from videos or audio inputs.
- Enriching the manual dictionary with the GloVe model.
- Calculate the LIWC.
- Analyze the results using ChatGPT to give an interpretation.

We chose some use cases to validate our work, which are presented in:

- Sentiment analysis.
- Political speech analysis.
- Mental health assessment.
- Customer feedback.

Our work distinguishes itself from previous similar work because it works on different kinds of multimedia, such as videos, audio recordings, and text. It is enriched by a pre-trained model, which is the GloVe model. It draws the word cloud. It also uses Large Language Models (LLMs) to do result interpretation.

In future work, we will enrich the dictionary more, try to cover all formats of textual data, add the Arabic language as another language for analysis, and integrate other LLMs like Gemini.

Bibliography

[1] الاتجاه نحو مقرر القياس النفسي وعلاقته بقلق الاختبار ودافعية الإنجاز لدى طلبة برنامج دبلوم الإرشاد الأسري بجامعة المجمعة. محمد زكري، علي. المجلة التربوية لكلية التربية بسوهاج.1455-1517:(2021)92.92 .

[2] اللغة وسيلة تواصل عبر الثقافات. الهنداوي هدى ، فتحي عبد المنصف محمد ، أبوشهبة منصور ، هاشم عجمي . مجلة بحوث كلية الآداب. جامعة المنوفية. (2019) ,377–391 ,118 , 30

[3] أداة. للتحليل اللغوي و حوسبة الكلمات باللغة العربية. زهو حياة، هابه مروة. جامعة الوادي (2023)

[4] Abburu, S., and Golla, S. B. Satellite image classification methods and techniques: A review. *International journal of computer applications 119*, 8 (2015).

[5] Andresen, B. B., and van den Brink, K. Multimedia in education. In *Infarmation technologies at school: conference materials* (2002), pp. 5–8.

[6] Bateson, M. C. *Arabic language handbook*, vol. 3. Georgetown University Press, 1967.

[7] Boyd, R. L., Ashokkumar, A., Seraj, S., and Pennebaker, J. W. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin* (2022), 1–47.

[8] Cai, L., Gao, J., and Zhao, D. A review of the application of deep learning in medical image classification and segmentation. *Annals of translational medicine 8*, 11 (2020).

[9] Camastra, F., and Vinciarelli, A. *Machine learning for audio, image and video analysis: theory and applications*. Springer, 2015.

[10] Chejne, A. G. *The Arabic language: Its role in history*. U of Minnesota Press, 1968.

[11] Chinchor, N. A., Thomas, J. J., Wong, P. C., Christel, M. G., and Ribarsky, W. Multimedia analysis+ visual analytics= multimedia analytics. *IEEE computer graphics and applications 30*, 5 (2010), 52–60.

[12] Ciani, A., Rosa, A., Santagata, R., et al. Video analysis as a learning tool to promote the quality of teaching: from school teachers' education to university teachers' professional development= la videoanalisi come dispositivo per promuovere la qualità della didattica: dalla formazione degli insegnanti allo sviluppo professionale dei docenti universitari. *GIORNALE ITALIANO DELLA RICERCA EDUCATIVA 14*, 27 (2021), 40–51.

[13] Culpeper, J. *History of English*. Routledge, 2015.

[14] Divakaran, A. *Multimedia content analysis: theory and applications*. Springer Science & Business Media, 2009.

[15] El-Kassas, W. S., Salama, C. R., Rafea, A. A., and Mohamed, H. K. Automatic text summarization: A comprehensive survey. *Expert systems with applications 165* (2021), 113679.

[16] Franco, M., Gaggi, O., and Palazzi, C. E. Analyzing the use of large language models for content moderation with chatgpt examples. In *Proceedings of the 3rd International Workshop on Open Challenges in Online Social Networks* (2023), pp. 1–8.

[17] Giannakopoulos, T., and Pikrakis, A. *Introduction to audio analysis: a MATLAB® approach*. Academic Press, 2014.

[18] Gong, S., Loy, C. C., and Xiang, T. Security and surveillance. *Visual analysis of humans: Looking at people* (2011), 455–472.

[19] Greene, S. S. *A grammar of the English language*. Cowperthwait & Company, 1867.

[20] Hao, S., Zhou, Y., and Guo, Y. A brief survey on semantic segmentation with deep learning. *Neurocomputing 406* (2020), 302–321.

[21] Islam, S., Elmekki, H., Elsebai, A., Bentahar, J., Drawel, N., Rjoub, G., and Pedrycz, W. A comprehensive survey on applications of transformers for deep learning tasks. *Expert Systems with Applications* (2023), 122666.

[22] Johnson, T. P. Myriad options in multimedia. *Landscape architecture 85*, 9 (1995), 26–31.

[23] Kelleher, J. D. *Deep learning*. MIT press, 2019.

[24] Kimball, L. G. *The structure of the English sentence*. American Book Company, 1900.

[25] Lerch, A. *An introduction to audio content analysis: Music Information Retrieval tasks and applications*. John Wiley & Sons, 2022.

[26] Ma, P., Li, C., Rahaman, M. M., Yao, Y., Zhang, J., Zou, S., Zhao, X., and Grzegorzek, M. A state-of-the-art survey of object detection techniques in microorganism image analysis: from classical methods to deep learning approaches. *Artificial Intelligence Review 56*, 2 (2023), 1627–1698.

[27] Mirończuk, M. M., and Protasiewicz, J. A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications 106* (2018), 36–54.

[28] Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. The development and psychometric properties of liwc2015. *Austin, TX: University of Texas at Austin* (2015).

[29] Rane, N., Choudhary, S., and Rane, J. Gemini versus chatgpt: Applications, performance, architecture, capabilities, and implementation. *Performance, Architecture, Capabilities, and Implementation (February 13, 2024)* (2024).

[30] Rezaeenour, J., Ahmadi, M., Jelodar, H., and Shahrooei, R. Systematic review of content analysis algorithms based on deep neural networks. *Multimedia Tools and Applications 82*, 12 (2023), 17879–17903.

[31] Singh, A., and Singh, P. Image classification: a survey. *Journal of Informatics Electrical and Electronics Engineering (JIEEE) 1*, 2 (2020), 1–9.

[32] Stevenson, A. *Oxford dictionary of English*. Oxford University Press, USA, 2010.

[33] Taboada, M. Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics 2* (2016), 325–347.

[34] Vidal, B. d. C. Image analysis and definition of nuclear phenotypes. *Brazilian Journal of Genetics 20* (1997), 749–754.

[35] Wang, J. R., and Parameswaran, N. Survey of sports video analysis: research issues and applications. In *Proceedings of the Pan-Sydney area workshop on Visual information processing* (2004), pp. 87–90.

[36] Wang, Y., Liu, Z., and Huang, J.-C. Multimedia content analysis-using both audio and visual clues. *IEEE signal processing magazine 17*, 6 (2000), 12–36.

[37] Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., and Zhang, Y. A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing* (2024), 100211.

[38] Yun, Y., and Kim, J. Cic: A framework for culturally-aware image captioning. *arXiv preprint arXiv:2402.05374* (2024).

[39] Zhou, P., Wang, L., Liu, Z., Hao, Y., Hui, P., Tarkoma, S., and Kangasharju, J. A survey on generative ai and llm for video generation, understanding, and streaming. *arXiv preprint arXiv:2404.16038* (2024).

[40] Павлов, Коваленко, and Голубкова. The basics of english syntax.