



UNIVERSITY OF WESTERN MACEDONIA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
&
UNIVERSITY OF WESTERN MACEDONIA
FACULTY OF SCIENCES
DEPARTMENT OF INFORMATICS
&
UNIVERSITY OF PIRAEUS
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
DEPARTMENT OF INFORMATICS

Conversational AI in Education: Use Cases, Challenges and Opportunities

DIPLOMA THESIS

of

DIMITRA-MICHAELA KOSETSIDOU

(AEM:228)

Supervisors : **Stergios Tegos** **Apostolos Mavridis**
Lecturer, UOWM Lecturer, UOWM

Kastoria December - 2023

Η παρούσα σελίδα σκοπίμως παραμένει λευκή



UNIVERSITY OF WESTERN MACEDONIA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
&
UNIVERSITY OF WESTERN MACEDONIA
FACULTY OF SCIENCES
DEPARTMENT OF INFORMATICS
&
UNIVERSITY OF PIRAEUS
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
DEPARTMENT OF INFORMATICS

Conversational AI in Education: Use Cases, Challenges and Opportunities

DIPLOMA THESIS

of

DIMITRA-MICHAELA KOSETSIDOU

(AEM: 228)

***Supervisors :* Stergios Tegos Apostolos Mavridis**
Lecturer, UOWM Lecturer, UOWM

Approved by the examination committee on 24 January 2024.

.....
Apostolos Mavridis
Lecturer, UOWM

.....
Angelos Michalas
Professor, UOWM

.....
Stergios Tegos
Lecturer, UOWM

Kastoria December – 2023

Copyright © 2023 – Dimitra-Michaela Kosetsidou

This work is copyright and may not be reproduced, stored nor distributed in whole or in part for commercial purposes. Permission is hereby granted to reproduce, store and distribute this work for non-profit, educational and research purposes, provided that this work and its corresponding publications are acknowledged.

The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the University of Western Macedonia..

I hereby declare that this thesis represents my original work and that I have used no other sources except as noted by citations.

Acknowledgments

I would like to express my gratitude to my advisors, Apostolos Mavridis and Stergios Tegos, for giving me the opportunity to explore the fascinating field of conversational AI and for their invaluable guidance throughout the realization of this thesis.

I would also like to thank the department staff for their assistance during my academic endeavor.

Περίληψη

Η αξιοποίηση των διαλογικών τεχνολογιών στην εκπαιδευτική διαδικασία αποτελεί αντικείμενο ερευνών εδώ και πολλά χρόνια. Ωστόσο, οι πρόσφατες τεχνολογικές εξελίξεις στο πεδίο της Τεχνητής Νοημοσύνης, σε συνδυασμό με την αυξανόμενη υιοθέτηση μεγάλων προ-εκπαιδευμένων γλωσσικών μοντέλων, όπως στην περίπτωση του πολυσυζητημένου ChatGPT της OpenAI, έχουν κεντρίσει την προσοχή ιδρυμάτων και επιχειρήσεων οδηγώντας σε έναν μακρύ διάλογο σχετικά με τις δυνατότητες και τα αποτελέσματα των εφαρμογών τους στον εκπαιδευτικό τομέα. Ο σκοπός της παρούσας μελέτης είναι διττός. Αρχικά, επιδιώκει να αναλύσει τα συστατικά στοιχεία πίσω από τον σχεδιασμό και την υλοποίηση των διαλογικών συστημάτων έως και σήμερα, παρακολουθώντας την εξελικτική πορεία των μηχανισμών αλλά και των δυνατοτήτων τους. Κατά δεύτερον, επιχειρεί να εξετάσει μέσα από μια σειρά περιπτώσεων αξιοποίησής τους στην εκπαιδευτική διαδικασία τα οφέλη που προσφέρουν στο σύγχρονο μαθησιακό περιβάλλον και τις προκλήσεις οι οποίες προκύπτουν από τη χρήση τους, παράλληλα με τις προτεινόμενες ενέργειες για την αντιμετώπιση αυτών των πτυχών. Τα πρώτα βήματα των υποσχόμενων αυτών τεχνολογιών στο χώρο της εκπαίδευσης συνοδεύονται από την ανάγκη για τη δημιουργία επαρκών προδιαγραφών, προκειμένου να ενσωματωθούν επιτυχώς στην εκπαιδευτική διαδικασία. Ωστόσο, η ανάδειξη μιας σειράς ευκαιριών προς την καινοτομία και τον μετασχηματισμό της μάθησης, της διδασκαλίας και της έρευνας που υποστηρίζουν και ωφελούν τα σύγχρονα εκπαιδευτικά περιβάλλοντα ανοίγουν το δρόμο προς τη συνέργεια των μελών της εκπαιδευτικής κοινότητας και τη συνέχιση της έρευνας, η οποία θα ξεδιπλώσει το σύνολο των δυνατοτήτων τους ως ένα ασφαλές και αποτελεσματικό εκπαιδευτικό εργαλείο.

Λέξεις Κλειδιά: Τεχνητή νοημοσύνη, Chatbot, Μεγάλο γλωσσικό μοντέλο, ChatGPT, ΤΠΕ, Αναδυόμενες τεχνολογίες, Εκπαίδευση

Abstract

The employment of conversational systems in the educational process has been a long-standing research topic. Nevertheless, the recent technological advancements in Artificial Intelligence, along with the increasing adoption of large pre-trained language models, such as OpenAI's much-discussed ChatGPT, have captured the attention of institutions and enterprises, leading to a long debate on the potential and results of their applications in the educational field. The purpose of this study is dual. First, it seeks to analyze the components behind the design and implementation of dialogue systems up to date, tracing the evolutionary path of their mechanisms but also of their capabilities. Secondly, it attempts to examine through a series of use cases in the educational process the benefits they offer in the contemporary learning environment and highlights the challenges that arise from their use, along with proposed actions to address these issues. The first steps of these promising technologies are accompanied by the need to create adequate specifications for their successful integration into education. However, the emergence of considerable opportunities for innovation and transformation of learning, teaching, and research that support and benefit modern educational environments pave the way towards synergy between members of the education community and continued research that will unleash their full potential as safe and effective educational tools.

Key Words: Conversational AI, Chatbot, Large language model, ChatGPT, ICT, Emergent technologies, Education

Contents

Introduction.....	13
1. Background.....	15
1.1 Conversational Artificial Intelligence	15
1.2 Conversational Agents/Chatbots	15
1.2.1 Historical Overview.....	16
1.2.2 Taxonomy.....	20
1.2.3 Early Implementation Approaches and Techniques.....	22
1.2.4 Architecture	28
2. Artificial Intelligence Approaches in Chatbot Development	31
2.1 Artificial Intelligence-Based Chatbots	31
2.2 Design and Implementation Techniques	32
2.2.1 Neural Network Models	32
2.2.2 Natural Language Processing.....	37
2.2.3 Language Modeling.....	38
2.3 Pre-Trained Language Models: Pre-training and Fine-tuning	41
2.3.1 Pre-Trained Transformer Language Models and Training Objectives.....	42
2.3.2 Fine-Tuned Language Models for Dialogue	43
2.4 Large Language Models: Scaling Pre-Trained Language Models.....	44
2.4.1 Scaling and the Emergent Abilities of Large Language Models.....	45
2.5 ChatGPT: The Case of a Prominent LLM-based Chatbot.....	46
2.5.1 OpenAI.....	46
2.5.2 GPT Architecture and Initial Models	47
2.5.3 GPT-3, 3.5, and 4 Model Background.....	48
2.5.4 Introduction to ChatGPT.....	49
2.5.5 Prompting and Prompt Engineering.....	51
2.5.1 Capabilities, Risks, and Limitations	53
3. Conversational AI In Education	56
3.1 Information and Communication Technologies and Emerging Trends in Education	56
3.1.1 Emerging Educational Technologies.....	57
3.2 Educational Chatbots.....	58
3.3 Use Cases of Chatbots in Education	60
3.3.1 Chatbots For Teaching Purposes	61
3.3.2 Language Learning	64
3.3.3 Chatbots for Assisting Purposes	66
3.3.4 Chatbots For Mentoring Purposes	69

3.4	Unleashing the potential of LLM-driven chatbots in education.....	72
3.4.1	Cases of Chatbot Prompts	73
3.5	Challenges	90
	Conclusions	94
	References	96

List of Figures

Figure 1: Chatbot systems through the years [21].	18
Figure 2: The Eliza chatbot [27].	21
Figure 3: Taxonomy of chatbots [13].	23
Figure 4: AIML code example [22].	25
Figure 5: ChatScript code example [13].	26
Figure 6: Universal chatbot architecture [13].	30
Figure 7: An unrolled recurrent neural network [30].	33
Figure 8: Components of an LSTM network [31].	34
Figure 9: A basic Seq2Seq model [34].	35
Figure 10: Transformer architecture [37].	36
Figure 11: Multi-head attention during the encoding of the token “it” [39].	37
Figure 12: Demonstration of a word embedding space [50].	41
Figure 13: The three classes of modern pre-trained models and their training objectives [42].	43
Figure 14: Timeline of Large Language Models [40].	46
Figure 15: Training process of GPT-3.5 models [69].	51
Figure 16: In-context learning and chain-of-thought-prompting strategies [40].	52
Figure 17: Custom instructions and generated output for a third-grade science lesson plan [73].	55
Figure 18: Learning Perspectives in the Education 4.0 context [87].	58
Figure 19: The RiPE principles for educational chatbots [4].	60
Figure 20: Interaction with the Ghost Storytelling Chatbot [98].	63
Figure 21: Top: a dialogue between a learner and the Elinor learning partner. Bottom: the design of Elinor’s background during discussions [100].	65
Figure 22: A student conversing with the Sister Fish learning companion [101].	66
Figure 23: AI technologies for the creation of a Jill Watson CA. [117]	69
Figure 24: Chat capabilities of Woebot [126].	71
Figure 25: A discussion between a user and LitBot about the user’s view of the learning material. It concludes with further recommendations from the chatbot based on the learner’s interests [125].	72
Figure 26: On the left, the conversation options are presented, as provided by the new Bing, along with the chat interface. On the right, the custom prompt preferences in terms of the subject, writing tone, format, and length.	74
Figure 27: ChatGPT’s introductory course as a Greek tutor, named GreGPT.	75
Figure 28: ChatGPT instructs cultural and vocabulary aspects of the Greek language (part 2).	75
Figure 29: ChatGPT as a Greek tutor recommending tips for language learners.	75
Figure 30: Bing Chat's display of an educational video about the Acropolis accompanied by video details. It constitutes the output of the prompt: “Show me an educational video about Acropolis”.	76
Figure 31: ChatGPT as a curriculum designer for a Greek language course for intermediate students (part 1).	77
Figure 32: ChatGPT as a curriculum designer for a Greek language course for intermediate students (part 2).	77
Figure 33: ChatGPT in the role of a grammar and spelling improver.	78
Figure 34: Bing Chat generating a gamified geography quiz for elementary school students.	79
Figure 35: Outputs of Bing Chat to correct and incorrect students' responses during a geography quiz, including rewards and inspirational feedback.	79

Figure 36: ChatGPT as a science lab creator, presenting general lab details.	80
Figure 37: ChatGPT analyzing a science experiment during a science lab simulation.....	80
Figure 38: ChatGPT provides educators with auxiliary information about the implementation of a successful science lab.	81
Figure 39: ChatGPT reenacting the prominent politician of modern Greek history, Eleftherios Venizelos, in the Greek language.	82
Figure 40: Bing Chat generates a fill-in-the-blank set of questions for a high-school biology class.....	82
Figure 41: ChatGPT as a creator of a multiple-choice assessment question for a high- school chemistry class.	84
Figure 42: ChatGPT as a creator of short-answer and matching assessment questions for a high-school chemistry class.	84
Figure 43: ChatGPT as a creator of essay and problem-solving assessment questions for a high-school chemistry class.	85
Figure 44: ChatGPT as a creator of critical thinking and application assessment questions for a high-school chemistry class.....	85
Figure 45: ChatGPT’s generation of an email invitation for an educational seminar (part 1).....	86
Figure 46: ChatGPT’s generation of an email invitation for an educational seminar (part 2).....	86
Figure 47: ChatGPT’s generation of an email invitation for an educational seminar (part 3).....	86
Figure 48: ChatGPT as a motivational coach.	87
Figure 49: ChatGPT as a meta-learning expert proposing strategies for learning the JavaScript programming language.	89
Figure 50: ChatGPT elaborates on the chunking meta-learning strategy based on JavaScript concepts.	90

List of Abbreviations

AI - Artificial Intelligence
AIML - Artificial Intelligent Markup Language
ALICE - Artificial Intelligent Internet Computer Entity
ANN - Artificial Neural Network
API - Application Programming Interface
AR - Augmented Reality
BERT - Bidirectional Encoder Representations from Transformers
CA - Conversational Agent
CoT - Chain-of-Thought
DL - Deep Learning
EC - Educational Chatbot
EFL - English as a Foreign Language
FAQ - Frequently Asked Questions
FERPA - Family Educational Rights and Privacy Act
GDPR - General Data Protection Regulation
GPT - Generative Pre-Trained Transformers
GRU - Gated Recurrent Unit
ICT - Information and Communication Technologies
IoT - Internet of Things
IR - Information Retrieval
IT - Information Technology
ITS - Intelligent Tutoring Systems
LLM - Large Language Model
LM - Language Model
LMS - Learning Management Systems
LSTM - Long Short-Term Memory Networks
ML - Machine Learning
MOOC - Massive Open Online Course
NLG - Natural Language Generation
NLP - Natural Language Processing
NLU - Natural Language Understanding
PLM - Pre-Trained Language Model
QA - Question and Answer
RL - Reinforcement Learning
RNN - Recurrent Neural Networks
Seq2Seq - Sequence to Sequence
VR - Virtual Reality
XML - Extensible Markup Language

Introduction

Recent years have seen the rapid growth and adoption of conversational artificial intelligence, attracting the interest of researchers. Conversational agents or chatbots that simulate and automate conversations with human users have penetrated various areas of our lives, bringing a wealth of new possibilities to the way we communicate, interact with computers, and handle our daily tasks.

In the educational context, previous research highlighted numerous benefits of these systems, emphasizing their ability to support individualized, adaptive, and interactive learning, as well as increase student engagement. However, studies indicate that despite increasing publications regarding the pedagogical use of chatbots continue to appear, research in the area remains underdeveloped [1], [2], [3], especially in real-world educational settings [4]. At the same time, efforts to shift from traditional to innovative learning environments, embracing emerging Information and Communication Technologies (ICT) to enhance learning outcomes and cultivate valuable 21st-century skills, give conversational artificial intelligence a dynamic and critical role that seems to unlock a range of pedagogical possibilities.

Additionally, the continuous advances in Artificial Intelligence (AI) methods and techniques, such as Natural Language Processing, especially after the rise of large language models, as in the case of OpenAI's much-discussed ChatGPT, raise myriads of questions and promises concerning the intriguing features, as well as the impact of this technology, creating a fertile research field concerning its educational utilization and implications that is presently at a nascent stage [5].

The purpose of this thesis is to provide, through extended literature, a comprehensive understanding of the multifaceted field of conversational AI, focusing on its educational applications. Furthermore, it endeavors to contribute to the ongoing discourse surrounding reformative and reinventing opportunities of this technology and drive further research that will unlock its full potential in the pedagogical process.

The thesis unfolds into three main chapters:

The first chapter lays the foundation by defining the core concepts of this study, namely conversational AI, conversational agents, and chatbots. It encompasses the

historical and technological evolution of the technology, offering insights into early development approaches, proposed taxonomies, and common architectural elements.

The second chapter delves into the AI components behind chatbot implementation, including Neural Network models and Natural Language Processing techniques, with a particular emphasis on the progress of language models. It encompasses the attributes of recent large-scale language models and their utilization for conversational situations, culminating with the presentation of ChatGPT, as a highly influential instantiation of an LLM-enabled chatbot, in which its technical and operational features are unraveled.

The third chapter examines the use of conversational AI technology in the realm of education. Initially, it provides a brief overview of ICT and emerging educational technologies, focusing on educational chatbots. Their pedagogical role and benefits are discussed and illustrated through numerous use cases. Finally, it underlines the potential and challenges of their incorporation into educational settings, along with potential mitigation tactics.

As a last note, the conclusions of the study are discussed, accompanied by suggestions for future directions.

1. Background

The current chapter introduces the term Conversational Artificial Intelligence and the systems that support this technology, namely conversational agents or chatbots. The main categories and architectural components of these systems are outlined, along with the methods initially employed in their development.

1.1 Conversational Artificial Intelligence

Conversational Artificial Intelligence is a sub-domain of AI that concerns the study of techniques involved with software agents capable of engaging in conversations with humans or other agents through natural language. These techniques include Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP), accompanied by dialogue management systems to understand and generate natural language [6]. Their application aims to optimize user communication by providing a speech or text-based interaction interface across various platforms, such as messaging services, websites, smartphones, web applications, or operating systems.

The versatile nature of conversational AI has emerged as an efficient solution for various application contexts. From healthcare, tourism, and educational applications to customer service, e-commerce, and entertainment, these agents have been extremely useful in facilitating communication between humans and computers. Combined with the fact that during the last decade, the increasing need for building more robust models has driven enormous progress in methods that are utilized for the development of these systems, it is not surprising that the demand for this technology, as well as research interest, is facing ever-expanding dimensions in popularity [7].

1.2 Conversational Agents/Chatbots

Conversational Agents (CAs) or Chatbots are systems enabling conversational AI technology. They constitute computer programs designed to communicate with humans in natural language via speech, text, or multimodal conversations [8] [9]. Depending on their capabilities, application areas, and type of embodiment [6], they appear with numerous terms in the research literature. Such terms are Conversational User Interfaces, Embodied Conversational Agents (ECAs), Virtual Personal

Assistants, Personal Digital Assistants, and Automatic Agents. Pedagogical Conversational Agents are a specific category of Pedagogical Agents or Intelligent Tutoring Systems (ITS), which are virtual characters intended to assist users in their learning process. They interact with learners in a conversational form, and their proposed design is based on social, emotional, cognitive, and pedagogical features [10] [11].

In recent academic literature, it has been observed that the terms Dialogue Systems and Chatbots are used as synonyms for Conversational Agents. However, there are also studies in which the terms are distinguished. Conversational agents, in particular, are mentioned as a category of dialogue systems [11]¹, and chatbots as a type of conversational agent [8] [12]².

1.2.1 Historical Overview

In 1950, Alan Turing posed whether machines can think, proposing the Turing Test, originally called the Imitation Game. The concept of this test involves an intelligent machine capable of misleading individuals into believing that they are interacting with another human being. The idea is regarded as a source of inspiration and is often deemed as the purpose of driving the development of the first and several subsequent chatbot technologies [13]. For an illustration of chatbots' history, see Figure 1.

The first generation of CAs began their appearance in the 1960s with text-based dialogue systems for question answering, as well as chatbots aiming to mimic human conversations [9]. ELIZA, the first chatbot implementation, was a stepping stone for the creation of chatbot technologies and promoted further field research [14]. It was built in 1966 by Weizenbaum to simulate a Rogerian³ psychotherapist. It used keyword identification and matching techniques in split input statements and returned

¹ According to Allouch et al. (2021), whereas dialogue systems use natural language to communicate with users, conversational agents are dialogue systems capable of understanding and generating natural language. The use of natural language does not strictly rely on a specific set of words or sentence structure, such as in Interactive Voice Response systems or voice command systems, which do not entail extended natural language content as input or output.

² McTear (2021) and Wahde and Virgolin (2022) refer to chatbots as non-task-oriented agents engaging in casual dialogues with users, as opposed to task-oriented agents aiming to accomplish specific tasks.

³ Rogerian psychology is a branch of clinical psychology where, during therapy, patients' words are reflected to them by the therapist [27].

an answer using a response mechanism based on templates. For instance, given the input string: I am sad, it takes ‘sad’ as a keyword and selects as an output: “How long have you been ‘sad’?” [15]. Despite its inflexibility, limitations regarding knowledge and context, as well as its support for short conversations [14], people were impressed by its capabilities that they trusted their private matters to the system [16]. A dialogue between a user and ELIZA is illustrated in Figure 2.

A few years later, in 1972, the rule-based chatbot PARRY was built by Kenneth Colby. It was implemented in the psychology domain, as well, to study patients’ behavior with schizophrenia. PARRY appeared to have “personality” traits, such as variables with values of fear and anger, which, affected by certain user sentences, would increment [16], thus leading to the expression of “emotional responses” [13]. Based on an experiment in 1979, where a sample of five psychiatrists were interviewed to determine whether they were conversing with a real patient or computer software, only one gave correct diagnoses. Although interview transcripts with PARRY closely resembled those with schizophrenic patients [16], the system suffered limitations regarding language comprehension and timely responses, in addition to its inability to convey emotions and learn from the discussion [13].

In the 1980s, speech-based dialogue systems developed by research laboratories within academia and industry followed, along with Voice User Interfaces (VUI) and Interactive Voice Response (IVR) systems, built simultaneously by enterprises for automated self-service tasks. Alongside the previous systems, ECAs also arose, equipped with virtual or physical bodies, as computer programs enriching the conversational aspect of interactions with body language [9].

Jabberwacky is a chatbot created by Rollo Carpenter in 1988 aimed to mimic a natural human conversation in an engaging and amusing approach [17]. It is written in CleverScript⁴ a spreadsheet-based language, where the identified inputs and the available outputs are placed on different lines of the spreadsheet, thus simplifying the development process. The system uses contextual pattern-matching techniques for its responses by accessing previous chats [13]. This task distinguishes the chatbot from its predecessors as it can learn and gradually refine its capacities [12]. However, it operated with a few users and came with speed limitations. In 2006, Cleverbot was

⁴ <https://www.cleverscript.com/about/>

launched as its successor, capable of holding more interactions and undergoing further adjustments [18].

In 1991, Michael L. Mauldin referred to a “Chatterbot”, a computer-controlled player in the TINY-MUD (multiplayer real-time virtual world), whose main goal is interacting with other players. It was a successful implementation since other players were not aware of the system’s presence, and they assumed all players were humans. They would only get suspicious if they came through a significant system mistake [19].

A year later, a chatbot called Dr. Sbaitso (an acronym for Sound Blaster Artificial Intelligent Text to Speech Operator) [13] was created for MS-DOS-based personal computers by CreativeLabs. It was designed to act as a psychologist by displaying the digitized voices produced by distributed sound cards. Most of its responses were standard (e.g., “WHY DO YOU FEEL THAT WAY?”) and did not involve any form of complicated interaction [20].

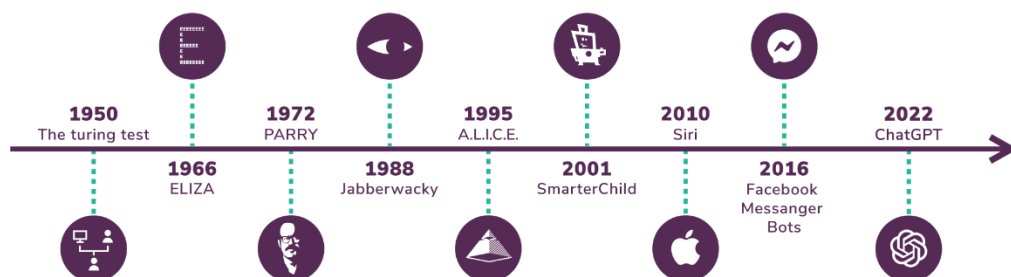


Figure 1: Chatbot systems through the years [21].

Another significant contribution towards chatbot technology was the development of A.L.I.C.E. (Artificial Intelligent Internet Computer Entity) in 1995, the first online chatbot inspired by ELIZA. A.L.I.C.E. used Artificial Intelligence Mark-up Language (AIML), an extension of XML, and pattern-matching to expand and enhance its software Knowledge Base (i.e., the set of manually written responses for users’ inputs) by adding new data objects. A.L.I.C.E utilized 41,000 templates and patterns, in contrast to ELIZA’s 200 keywords and rules [13]. The system won three times the

Loebner Prize Competition⁵, an annual implementation of the Turing Test, and has been ranked as the “most human computer” [22]. Among its characteristics is enabling longer and broader conversations online but lacking human-like responses.

An evolutionary stage in chatbot technology was established in 2001 with the release of SmarterChild on America Online (AOL) and MSN Messenger. It was the first endeavor to create a personal assistant for practical purposes (e.g., calculations, information about the weather, and news), able to obtain database information through user queries [13].

The technological progress in AI, particularly in NLP, has driven the creation of new generations of CAs [8]. Voice assistants (also known as smart personal assistants), a successor of SmarterChild, paved the way. Commencing from the launch of Apple’s Siri in 2010, a personal assistant capable of interacting with users and helping perform tasks through the iPhone 4S, the continuous integration of conversational agents operating on various platforms and devices (e.g., smartphones, smart speakers, and cars) led to the processing of compound natural language [23] and provided answers to more complex user queries. With IBM’s Watson in 2011, Google Now in 2012 (a prior version of 2019 Google Assistant), Microsoft’s Cortana, and Amazon’s Alexa released in 2014, voice assistants began acquiring growing popularity. Their innovative elements allow them to learn through a variety of data and adapt to various areas (e.g., health, education, and customer service), providing a more personalized user experience and enhanced human-computer interaction. By employing DL algorithms and being integrated into smart devices, they can be activated by a wake word (voice command) [14], connect to the Internet, swiftly generate meaningful answers, understand and execute tasks (e.g., controlling home devices, accessing calendars, making phone calls).

During 2016, there was a notable increase in the use of chatbots in both academic and industrial settings. Thousands of chatbots were built primarily for customer service on social media platforms. Subsequently, they were utilized for a variety of research and operational purposes. The simultaneous advent of the Internet of Things (IoT) was beneficial since it allowed for seamless communication with other connected smart devices [13].

⁵ https://en.wikipedia.org/wiki/Loebner_Prize

Along with chatbots assisting users to complete daily tasks, social chatbots with entertainment purposes began appearing. Microsoft XiaoIce [24] is an AI chatbot intended to provide a long-term companion and socialize with human users [14] [13]. Designed with personality traits, it holds Intelligent Quotient (IQ), such as knowledge and memory modeling, reasoning, image and natural language comprehension, and Emotional Quotient (EQ), taking into consideration cultural differences and ethical issues. It also features both information-retrieval and generative-response ability, enabling extended and open-domain conversations with users.

In the last few years, open-domain chatbots have ushered in a new era due to recent advances in DL techniques that are used to train them from large datasets of dialogues [8] [14]. Google's Meena is a generative chatbot based on a deep neural network architecture introduced in 2020, which outperformed modern chatbots in providing sensible replies accompanied by specific rather than unclear information [12]. Facebook's BlenderBot is another open-domain and open-source system enabling further user experimentation. It was presented the same year by the research team, which stated that the chatbot holds a versatile set of skills (e.g., personality and empathy traits) during interactions along with an improved response generation mechanism.

Currently, the transformer deep learning architecture-based language models, which follow pre-training and fine-tuning techniques, are regarded as the standard approach for conversational tasks. These technologies are behind industry-leading chatbots such as Facebook's BlenderBot, Open AI's ChatGPT, Google's Bard, Baidu's ErnieBot, and Microsoft's Bing Chat (also known as the new Bing).

1.2.2 Taxonomy

Several aspects have been taken into consideration within academia for the classification of CAs/chatbots, with the wide scope surrounding their use allowing them to fit into more than one category [13], [25]. Typical examples include their provided services, their permissions, the medium of interaction [11], and the human aid, as shown in Figure 3. This work will focus on classifications based on their knowledge domain, goals and applications, and response generation method.

Typically, the knowledge domain of a chatbot includes the range of knowledge it can access [13]. Accordingly, chatbots differentiate between **closed-domain**, which

can only respond to user questions from a particular domain as they are designed to achieve specific goals [26], and **open-domain chatbots**, capable of answering questions concerning multiple domains. Generic chatbots have also been referred to as systems that can answer any user question from every domain possible.

```
Welcome to
EEEEEE LL      IIII  ZZZZZZ  AAAAA
EE      LL      II   ZZ      AA  AA
EEEEEE LL      II   ZZZ      AAAAAA
EE      LL      II   ZZ      AA  AA
EEEEEE LLLLLL IIII ZZZZZZ  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Figure 2: The Eliza chatbot [27].

Depending on the goal a chatbot aims to achieve, it can be categorized as **task-oriented** (also referred to as goal-oriented) or **non-task-oriented** [8], [12]. Task-oriented chatbots are designed for conversations with users to accomplish particular tasks within one or more domains. Their ultimate goal is to answer users' queries with precise, relevant, and valid responses and to help them perform serious and complex tasks [12]. Voice assistants (e.g., Apple's Siri, Google Assistant/Home, Amazon's Alexa, Microsoft's Cortana, etc.) employ task-oriented agents to schedule meetings, provide weather information, or order items online. Various businesses employ task-oriented agents as customer support or sales assistants, while these agents have been adopted widely in education and healthcare settings.

On the other hand, non-task-oriented agents are intended to mimic human-like interactions, so instead of primarily focusing on the answer's details and precision, their goal is to maintain a more extended, informal dialogue (also known as chit-chat) surrounding everyday topics [12]. They have been applied to various fields, including education, business, e-commerce [8], and entertainment, allowing for solving tasks through a more natural and engaging dialogue [16].

Other researchers have proposed a similar distinction between task-based chatbots that undertake a more transactional function, such as flight booking, chat-based or conversational chatbots that focus on handling a natural cross-domain discussion with users (chit-chat), usually for entertainment [25], and informative chatbots that provide information to users, stored in a fixed source, such as FAQ chatbots [13].

Finally, the response generation methods used, typically connected to the design approaches followed, broadly categorize chatbots as either **rule-based** systems (also known as flow-based or pattern-based) that operate given a predefined set of rules or **AI-based** systems that use ML techniques, including corpus-based (also known as statistical data-driven) and end-to-end neural development systems. Corpus-based systems are built with enormous amounts of specific conversation datasets to extract human conversations [16]. Moreover, data stems from interactions with the chatbot by real or simulated users and is used for optimizing the chatbot's dialogue policy [8]. The last relies on a neural, end-to-end architecture in which the various components of the typical chatbot architecture, as presented in 1.2.4, are optional since there is a direct mapping from the input to the output, without the need for intermediate processing. AI-based chatbots are further categorized as Information-Retrieval and Generative, depending on their answering process. For more details, see section 2.1.

Hybrid systems have also surfaced by merging both approaches since nothing prevents their methods from being used together [12].

1.2.3 Early Implementation Approaches and Techniques

In the early days of chatbot development, rule-based systems were used to produce output. ELIZA, ALICE, PARRY, Chatterbot in "TinyMUD", and Jabberwacky are a few examples of such systems [13] using both simple as well as more sophisticated pattern-matching rules that take into account the conversational context to match a pattern (i.e., ALICE) [25]. Their simplicity in terms of building [14] is an advantage that often leads to their adoption by modern QA chatbots [15]. They operate well on closed-domain discussions as they restrict the responses to specific commands available in the chatbot's dataset. However, they cannot answer more complex user questions, especially those that do not align with the included keywords in the dataset [14], [28].

Rule-based model responses rely on pattern-matching algorithms. The user input matches a rule pattern. Afterward, an answer is chosen from a predefined set. New responses cannot be generated as the database of the ones available is composed of conversational patterns [13]. Manually crafted pairs are created that contain a <pattern> and a <template>. The former is a classification of the input, while the latter constitutes the database of the mentioned and most suitable answers [25]. Any unknown pattern from a sentence can lead to doubtful replies [14]. So, a more extended rule database leads to more possibilities for relevant and accurate answers [13].

While pattern-matching rules tend to respond quickly to user inputs due to a lesser deep semantic or syntactic processing, and they easily create conversations, the responses could lack originality and be repetitive, domain-dependent, and inflexible as they struggle to transfer from one context to another [14]. Additionally, the process of manually encoding these rules is often time-consuming and tedious, especially when high volumes of data need to be typed. Grammar or syntax mistakes in user answers are also challenging to manage [13].

Chatbot Categories	Knowledge domain	Generic
		Open Domain
		Closed Domain
	Service provided	Interpersonal
		Intrapersonal
		Inter-agent
	Goals	Informative
		Chat based/Conversational
		Task based
	Response Generation Method	Rule based
	Retrieval based	
	Generative	
Human-aid	Human-mediated	
	Autonomous	
Permissions	Open-source	
	Commercial	
Communication channel	Text	
	Voice	
	Image	

Figure 3: Taxonomy of chatbots [13].

1.2.3.1 Artificial Intelligence Markup Language

One of the most common development tools for rule-based CAs is the Artificial Intelligence Markup Language (AIML), developed for the creation of the Knowledge

Base of early chatbots, such as ALICE. It was also used for building the Mitsuku⁶ chatbot in 2005 by Stephen Worswik, which won the Loebner Prize Competition five times holding a world record [15].

It is an open-source language derived from XML, helpful for designing dialogue flow [25], easy to learn and use, and with available pre-written AIML collections for interested developers. It is often accompanied by other techniques, such as Latent Semantic Analysis (LSA)⁷, as many of the languages used in pattern-matching to produce replies cannot occur from templates [13].

AIML is a set of data objects called AIML objects. These objects are composed of topics, which are upper-level elements with a name attribute. Each topic contains a group of relevant categories representing knowledge units (i.e., dialogue). A category, represented by the <category> tag, encompasses a chatbot rule for matching input and output [25]. It contains at least a “pattern” (tag <pattern>) for a user’s input, including words, single spaces, and wildcard characters that replace a string or a part of the string (i.e., ‘_’ and ‘*’) as well as a “template” (tag <template>) for the chatbot’s response. A <srail > tag can be practical for repetitive mentions in the discussion and deliver a more natural tone, as it can enter a reply on its own as input to itself [15]. For a code snippet of AIML, see Figure 4.

All AIML patterns are stored in trees where the nodes represent the categories, and the leaves represent the templates [13]. Graphmaster, a wieldy tool responsible for the execution of the pattern-matching algorithm, manages these trees [15].

1.2.3.2 Chatscript

Chatscript is another rule-based chatbot development tool that uses pattern-matching. It features a built-in tagger and parser to analyze and modify user input. The tool uses concepts, which are groups of words that are relevant in meaning or other parts of speech, such as nouns and adverbs. Developers can easily access these concepts for chatbot building since they are stored in text files included in a database. Moreover, it is equipped with both short-term and long-term memory to handle

⁶ <http://www.square-bear.co.uk/mitsuku/home.htm>

⁷ A method of extracting semantic (meaning) similarities between words based on statistical computations on a document collection and vector representation [134].

variables that store user information. These variables can be used directly or in conjunction with other conditionals to formulate responses by the chatbot. It is also case-sensitive, which is helpful for detecting emotions in user replies. The Mitsuku, chatbot employs Chatscript alongside AIML [13]. You can see an example of Chatscript code in Figure 5.

```
<aiml version="1.0.1" encoding="UTF-8"?>
  <category>
    <pattern> My name is * and I am * years old </pattern>
    <template> Hello <star/>. I am also <star index="2"/> years old!</template>
  </category>
</aiml>
```

Figure 4: AIML code example [22].

1.2.3.3 Parsing

Parsing is the process of deconstructing an input string, such as a sentence, to analyze it grammatically and determine whether it applies to grammar rules [8]. The input text is categorized into noun and verb phrases with respective parts of speech, which can be easily stored and managed [25]. This process constitutes lexical parsing and was initially used in ELIZA to identify keywords and match them with corpus responses. This approach enables a chatbot, which has sparse templates and patterns, to generate responses for different input strings [15].

Semantic parsing, on the other hand, defines the meaning of the input sentence. The result can be displayed as a parse tree (similar to a syntax tree), demonstrating the relationships between the words and phrases involved [8]. These parsers are used in some commercial chatbot frameworks, such as Dialogflow, for intent identification [25].

1.2.3.4 Markov Chain Models

A Markov Chain Model mathematically refers to a model that expresses the probability of current events based on the state of past events. It is a probabilistic model that has been widely used for chatbots intended for entertainment and emulating simple human discussions [25].

In chatbot development, it seeks the probability of how many times a letter or a word appears in a dataset and utilizes this probability distribution to select the most appropriate sequences for the answers. Markov chains have orders that define how many letters or words will be chosen as input [15]. For a 0-order Markov chain in a string with just letters, such as “lanlnlaann”, the letter “l” appears with a probability of 2/10, the “a” with 3/10, and the “n” with 5/10. For order 1 Markov chain, it will take into consideration the precedent component for the calculation of the fixed probabilities. For order 2 Markov chain, given a sentence such as “the yellow bird fell off the tree”, the element “the yellow” will be followed by “bird”, the “yellow bird” will be followed by “fell” and the remaining words will comply in the same manner. If different inputs appear in two results, a probability of 0.5 will be employed in both input strings. Taking into account the present state, a possibility of one or more states follows that a chatbot can use to generate various, more relevant, and appropriate responses with a mutual degree of consistency [25].

Despite the convenience of programming and the ability to represent the model in matrix form, it has been observed that a Markov chain model faces significant challenges in attaining sophisticated dialogues [25].

```
concept:~greeting [hello hi hey]

concept:~goodnight [night goodnight ]
Topic:~chitchat( ~greeting)
t: ~greeting

#! Hello
u: ( * ~greeting * ) [Hi!] [Hello], [What's your name?] [How are you?]

#! My name is
u: ( * name * is _ * ) Nice to meet you _0!
$usename = '_0'

#! I wish you a goodnight
u: ( <<!not * ~goodnight * >> ) Goodnight $usename
```

Figure 5: ChatScript code example [13].

1.2.3.5 *Ontologies/Semantic Nets*

In computer science, ontologies are the formal representations of concepts (also called classes) of entities in a domain. Their properties (often called slots or roles) include different characteristics of concepts and slot restrictions (often called facets or role restrictions) [29]. In the field of medicine, for instance, an ontology would

represent knowledge about symptoms, diseases, and treatments [8]. They are a commonly used framework in artificial intelligence and the semantic web since they provide a structured form of organizing information.

Classes define concepts in a domain. For instance, a class of coffee refers to every kind of coffee, while individual kinds of coffee are instances of this class. The class can create subclasses, such as filter and instant coffee, where they jointly create a hierarchy graph. These subclasses form a coffee superclass. Slots of each class (which define the classes' properties and instances [29]) may incorporate the company of coffee or its grind. Slot restrictions could entail, for example, the value type and the type of packaging. The development of an ontology, consisting of different instances of classes, slot values, and facets, serves as a foundation for constructing a Knowledge Base [15].

Ontologies are a helpful tool for searching through the taxonomical hierarchies for relations between concepts in dialogues and can also lead to new reasoning and answer possibilities [25], [15]. In chatbots, OpenCyc and Wordnet ontologies have been employed to supplement their knowledge base. By updating the conversation state and searching the nodes of the Knowledge Graph⁸, the chatbot creates connections for the concepts used in the conversation [13].

1.2.3.6 Relational Database and Language Tricks

To enhance the chatbot's responding ability and form a more meaningful dialogue, [13] its Knowledge Base is enriched with a Relational Database (RDB) for accessing former conversations and aiding in the formulation of various responses to the same questions posed from time to time [15]. In this approach, Structured Query Language (SQL) seems to be the most frequently used language.

In addition, certain language techniques may serve as a means to perform human-like routines in conversations, such as mimicking typing errors, fake keystrokes or backward blanks, and illogical responses, as well as introducing standardized answers

⁸ Often, ontologies are used interchangeably with the term knowledge graph, but while a knowledge graph represents relationships between specific entities in one or more domains, an ontology models general entities and their relationships. In other words, a knowledge graph is an ontology that includes data [8].

and personality characteristics (e.g., date of birth, age, or parents) to the chatbot [13], [15].

1.2.4 Architecture

The proposed dialogue architecture is followed by the paradigm of [13] and concentrates on the primary features of both rule-based and AI chatbots, as illustrated in Figure 6.

1. User Interface

The chatbot's activity begins with a dialogue act⁹ (i.e., a user's query, command, or answer to a prompt), formed as speech or text input in an application, such as a messenger application (e.g., Facebook, WhatsApp, or Skype) [22].

In the case of a voice-based system, the dialogue act is delivered as an acoustic signal, which later passes as an input to the Automatic Speech Recognition component.

2. User Message Analysis

Next, the User Message Analysis component receives the input [13]. It utilizes either rule-based or machine-learning approaches to interpret the user intent (= intention) along with certain aspects (also known as entities) that need to be captured and extracted. The grammatical and syntactical features of this input might stay intact or be modified through NLP.

There are also cognitive services connected to the User Message Analysis that can offer better results, such as a spell checker to correct any mistakes in user inputs and help understand the user intent, a machine translator for multilingual chatbots, or sentiment analysis, especially cloud-based, to identify any positive or negative statements of users [13].

3. Dialogue Manager

The Dialogue Manager is responsible for maintaining the intent and entities at that stage, as well as monitoring and keeping the conversation context updated. It

⁹ Dialogue or speech acts represent actions as expressed in a user's utterance [16], such as question, command, promise, or threat form as a label [8].

comprises Dialogue State Tracking, which is also called Belief State Tracking or Dialogue modeling, and the Dialogue Policy or Dialogue control components [11], [8]. Dialogue State Tracking contains all the relevant information from the interactions to determine the next system's actions and updates the dialogue state when new information arises. Dialogue Policy determines the actions that are associated with the dialogue states [8]. The chatbot can ask the user for further input to complement any omitted context data or needed entities.

It involves ambiguity, data, and error-handling elements. The former undertakes to respond appropriately in the case of intent misunderstanding. That includes generic answers, requests for explanatory information, or even the unavailability of a reply. Data handling saves user data in a file to alter and personalize its replies, while error handling, as the name suggests, deals with unforeseen errors [13].

Once the intent is determined, the chatbot proceeds to deliver the desired results. It can either retrieve information from various data sources from the Backend or generate a response. In the first case, the Dialogue Manager retains control of the conversation flow, while in the latter case, the Backend undertakes this task [13].

4. Backend

The chatbot executes the demanded actions. Information retrieval commences through API requests or database queries [13]. After acquiring the necessary insight, it passes it to the Dialogue Manager and, ultimately, to the Response Generator.

5. Response Generator

The response is prepared based on the intent and context information returned from the user message analysis component [22]. Typically, the answering process is set by either rule-based implementation approaches or AI (i.e., IR-based and Generative-based models). Rule-based chatbots choose a response between a predefined set on the Knowledge Base, transferred directly by the Dialogue Manager. In IR-based models, the most suitable reply is selected from the available set of sources or through APIs, and in generative models, NLG takes place to give a more human-like answer in terms of the inputs preceded. In hybrid approaches, the selection is based on comparing the retrieved and generated replies. After responding to users, the chatbot waits for the user's feedback.

Within a voice-based system, the resultant output is rendered as an acoustic signal by the Text-to-Speech Synthesis component. This auditory signal may elicit additional input from the user, thus initiating another cycle in the process [8].

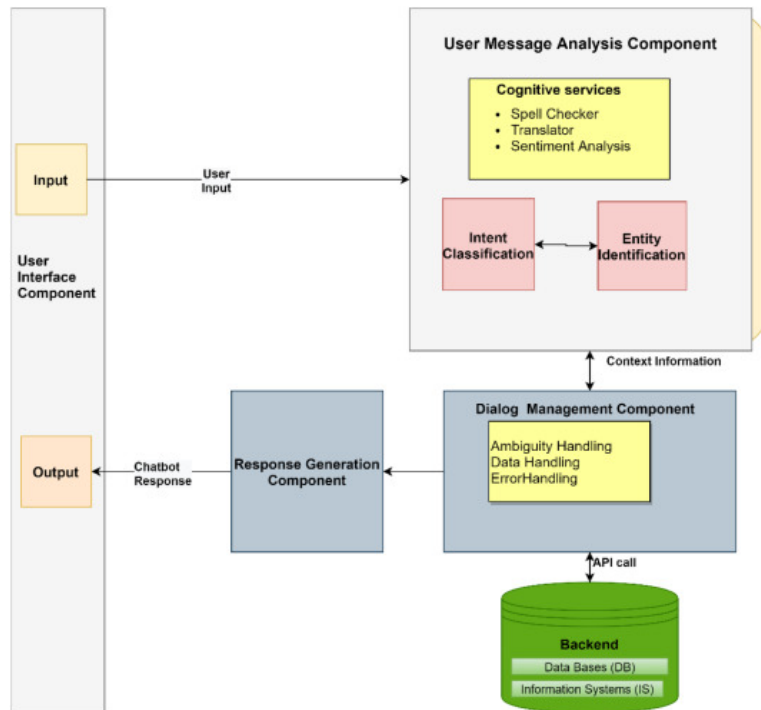


Figure 6: Universal chatbot architecture [13].

2. Artificial Intelligence Approaches in Chatbot Development

This chapter discusses the AI approaches utilized for the design and development of chatbots, exploring the area of NLP and covering the characteristics and advancements of language models. The circumstances surrounding the launch of the popular LLM-driven chatbot, ChatGPT, are also unfolded along with its capabilities and limitations.

2.1 Artificial Intelligence-Based Chatbots

Opposite to rule-based systems, AI-based chatbots employ ML models and NLP methods to acquire and analyze user input and produce replies. These models do not require pre-existing responses for each user statement, exact pattern matching, or coding of new rules. Instead, they are trained on extensive datasets of conversational data using ML algorithms that can learn from previous discourse and gradually become more intelligent [28]. These approaches contribute to more flexible systems that allow for a broad scope of knowledge [14]. Advantageously, the entire conversational context is considered rather than the current utterance each time [13]. AI chatbots are broadly divided into Information-retrieval and Generative chatbots, although there are hybrid examples of merging the two methods [15].

Information-retrieval (IR) chatbots aim to retrieve the necessary information through an existing dataset containing the available answers [14]. After user input processing, they select the most appropriate response on the set. Usually, the knowledge base of these chatbots consists of a QA-pair dataset, and it formulates a chat index that lists the potential responses based on the user utterance. User inputs or queries align with the most relevant in the chat index, and the chatbot returns the appropriate answer. An upsurge in the popularity of these systems arose due to the proliferation of Web 2.0 technologies and the consequent availability of large amounts of online textual data [15].

Since this approach contains pre-written responses, the quality is guaranteed [14]. However, synthesizing the knowledge base of these models and training them with large amounts of conversational data may be challenging and tedious. In addition, it may restrain the implementation of personality traits in social chatbots due to the inability to create responses.

Generative chatbots, on the contrary, typically rely on deep neural networks (e.g., Encoder/Decoder) and generate new answers from scratch to queries formulated from user input [14]. In the last few years, these systems have concentrated the focus of researchers, as they have shown themselves to be more suitable in open-domain dialogue. Generally, they are trained on an extended dataset of sentences derived from human conversations that algorithms use so that the model can learn to synthesize linguistically coherent, correct, and relevant responses.

2.2 Design and Implementation Techniques

2.2.1 Neural Network Models

Artificial Neural Networks (ANNs) are ML and DL models adopted in IR-based and Generative-based chatbot implementation approaches. Trained on large labeled or unlabeled datasets, they can model the relationships and patterns between input and output natural language data and produce the relevant answers.

In IR-based chatbots, the reply derives from probability computations on the most suitable response in the neural network [15]. In generative models, answers are produced one word at a time after probability calculations are applied to the available vocabulary. In hybrid approaches, a comparison occurs between the candidate responses, and the one ranked higher gets selected.

2.2.1.1 Recurrent Neural Networks

Deep neural networks have been widely applied in conversational modeling, starting from Recurrent Neural Networks (RNNs). The design of RNNs provides helpful assistance in handling the sequential natural language data and preserving context. Both features are valuable for more meaningful output [25].

Derived from ANNs and modified from recursive neural networks, an RNN is used for remembering previous input sequences of variable lengths and passing them to neurons [13], as opposed to a regular feed-forward neural network. The hidden layers of RNNs perform a **loop**, which enables an internal memory of knowledge related to former inputs and transferred throughout the network that can be used for further processing. By storing one layer's output and passing it to another as input,

RNNs can save all previous data and use it in current computations. Moreover, they can gather data related to the order of the input sequence that may be helpful for its total processing [8]. For an illustration of an RNN model, see Figure 7.

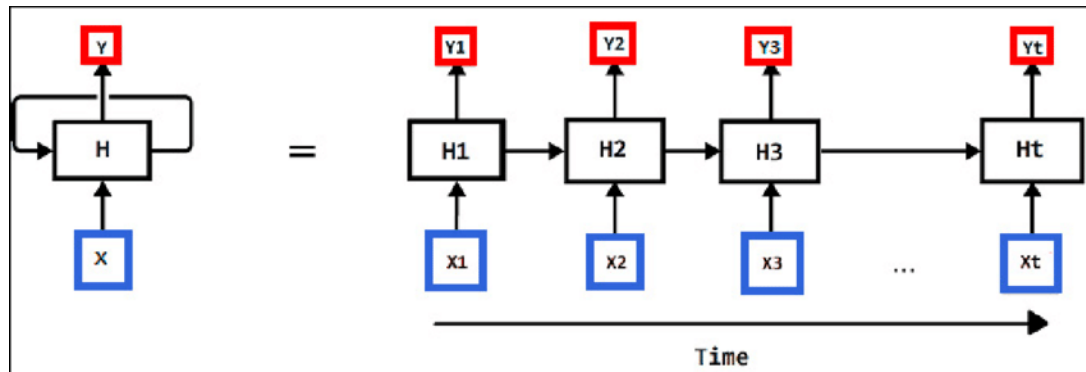


Figure 7: An unrolled recurrent neural network [30].

2.2.1.2 Long Short-Term Memory Networks

Searching the information history of a conversation is challenging for an RNN due to the issue of vanishing gradients, which complicates the learning of long-term dependencies. Long Short-Term Memory Networks (LSTMs) are a specific type of RNN formulated to overcome this difficulty and retrieve knowledge mentioned back in many prior steps [13] by supplying the network with a mechanism that controls the flow of information. This involves the integration of a complementary context layer to the network, the *cell state* [8]. The cell state in LSTMs includes “gates” that control the transfer of knowledge to the cell state and the subsequent network units by assigning a value between 0 and 1. A value of 0 means that nothing passes, while a value of 1 means that everything passes. The input gate determines the knowledge that needs to be preserved, while the forget gate resolves what needs to be removed, thereby keeping the network status updated. The output gate sets the output from the hidden layer [15]. For an illustration of an LSTM network, see Figure 8.

LSTM undertakes, for example, to recall the gender of the subject mentioned in an earlier input to choose the appropriate personal or demonstrative pronouns [15]. Addressing these issues allows the network to achieve better performance. It has often been utilized in chatbot design with other LSTM modifications, such as Gated Recurrent Unit (GRU), where a single “update gate” is created by merging the input and forget gate.

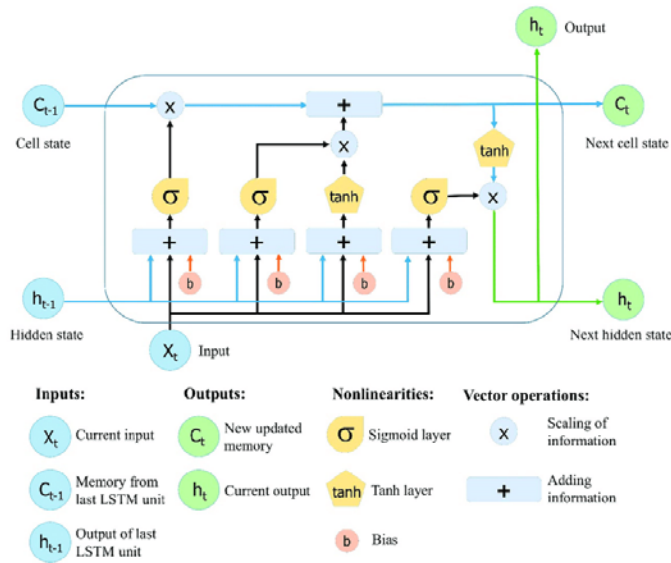


Figure 8: Components of an LSTM network [31].

2.2.1.3 Sequence to Sequence (Seq2Seq) Models

A Sequence to Sequence (Seq2Seq) neural model is a flexible, end-to-end, generative model developed primarily for machine translation tasks (the input sentence in the source language and the translated output in the target language). However, it appeared to be an efficient practice for modeling conversations (e.g., response generators of HappySoul and XiaoIce chatbots [32], [33]) and becoming a standard approach for modern NLP tasks [14].

The model is composed of two recurrent neural networks, including LSTMs and GRUs for modeling longer sentences [13], or Transformer networks; an **encoder**, on the one hand, that produces a vector for the input sequence or <status> processing, and a **decoder** for decoding the <status>, taking the vector and generate the preferred output or <response> [8] [25]. The intention behind the implementation of this model is to produce the most probable answer, considering the context as formed by the preceding turn or input sentence [14]. Initially, the input string is fed to the former RNN encoder and gets processed one word at a time in a hidden state of the network. The final state of the encoder reveals the context vector, the representation of the input sentence, which is then inputted into the decoder to generate an output, one element at a time, based on a proper probability function. An illustration is shown in Figure 9.

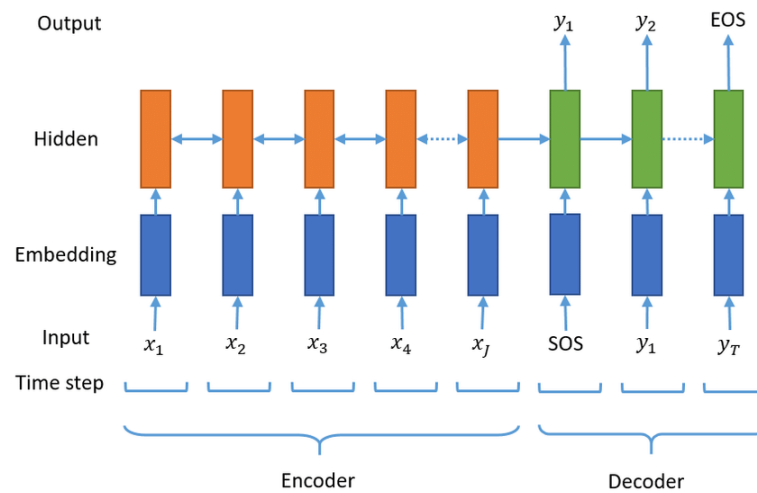


Figure 9: A basic Seq2Seq model [34].

2.2.1.4 Deep Seq2Seq Models

Seq2Seq models can be enhanced with multiple LSTM networks to boost their performance, forming Deep Seq2Seq models. These practices are useful for generative chatbot efficiency and assist them in acquiring more human-like characteristics [13].

During the design process of a Deep Seq2Seq model, the output is continuously promoted from the former layers to the subsequent [15]. The input sequence is supplied to the first encoder layer. Then, the LSTM encoder undertakes to transform all words into a vector. The output passes to the next LSTM encoder layer until the output derived from the final LSTM encoder layer routes to the first LSTM decoder layer. The process finishes with the target sequence calculated by the corresponding probability function.

2.2.1.5 Transformers

One of the limitations of Seq2Seq models is the fact that the whole input sequence needs to be encoded in a fixed-length context vector, so their performance decreases when they have to face longer sequences, leading to ill-defined answers [14]. The attention mechanism was proposed as assistance in this issue, where through mathematical techniques, the most relevant portions of the input sequence are selected by the decoder for the prediction of every subsequent word in the output [8].

In 2017, a Google team proposed in the paper “Attention is All You Need” the Transformer deep neural network structure [35], composed of an encoder and a decoder with similar layers for processing sequential data such as natural language, relying on attention mechanisms. Its key element, called *self-attention*, lies in the differentiated way in which each part of the input data is weighted considering its relevance to a particular word, enabling the detection of long-term dependencies and the contextual relationships within the sequence [14]. It permits parallel processing of the input sequence, thereby leading to model training with larger datasets at a faster rate. For an illustration of the Transformer architecture, see Figure 10.

Specifically, every sequence component has its own representation, and to estimate the relationship between these components, computations of the Q (*query*), K (*key*), and V (*value*) matrices are needed, which are produced from linear representations of the input sequence. The *query* matrix refers to the present component, the *key* matrix to other components, and the *value* matrix incorporates information that needs to be gathered. The computation of similarity between the query and key matrices defines the correlation weight of the current item with the others through a dot-product calculation. Afterward, the *softmax* function is performed to normalize the similarity so that the sum of each correlation equals 1. The new weights are assigned to the respective values that later accumulate, leading to a unified representation that includes information regarding the relationship of the current word with the others [36].

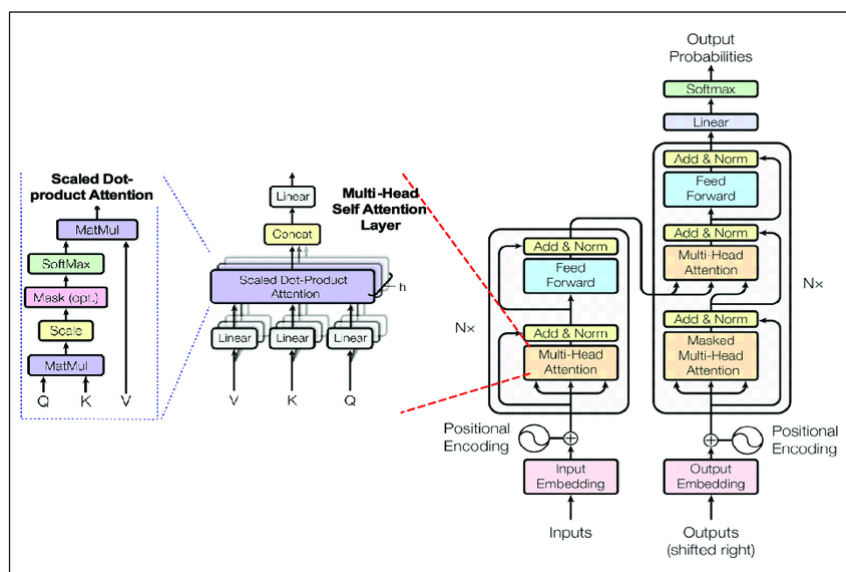


Figure 10: Transformer architecture [37].

Multi-head self-attention in a model is the result of the multiple and parallel applications of the attention mechanism (i.e., “heads”), each of whom has its learnable parameters and can focus on diverse data information, such as semantic or syntactic relationships (see Figure 11). The insight gathered by every head can aggregate into a single representation to form the desired output [38].

These powerful features introduced for neural machine translation proved profitable for numerous NLP tasks, establishing a “foundation” architecture with various implementations. Notable cases include Pre-Trained Language Models, such as Google’s Bidirectional Encoder Representations from Transformers (BERT) and OpenAI’s Generative Pre-Trained Transformers (GPTs).

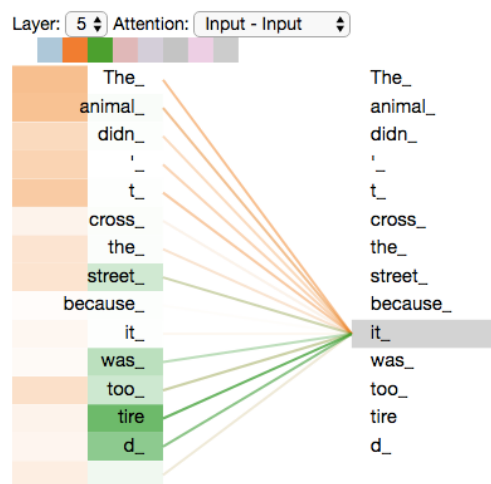


Figure 11: Multi-head attention during the encoding of the token “it” [39].

2.2.2 Natural Language Processing

For many decades, human ability to communicate has been the focus of AI research. The sophisticated nature of human language, in particular, has led to the development of multiple powerful AI algorithms and techniques attempting to assist machines in coping with understanding and handling natural language expressions. Natural Language Processing (NLP) is a branch of AI that investigates how computer systems handle natural language speech or text [22]. This field has yielded valuable insights into language technology, enabling the creation of techniques that facilitate the effective manipulation of human language and the execution of intended actions.

Typically, these practices are based on ML encompassing NLU to comprehend natural language content and NLG that undertakes to produce natural language output that is meaningful, non-repetitive, complete, and precise [28].

NLU components are fundamental for implementing natural user interfaces, such as chatbots [22]. They try to recover context and meaningful information from the user's unstructured utterances and give a suitable reply given the user's intention. They underlay tasks, such as text tokenization (the process of decomposing a given text into segments called tokens, which consist of words, phrases, or punctuation marks), parsing [11], intent detection and classification, and entity extraction, considering the contextual background [13].

Three concepts need to be clarified at this point: intent, entity, and contexts. *Intent* stands for the representation of the user's intention. It is a connection between the user's input and the steps that need to be carried out by the chatbot which may include parameters to determine more detailed feedback [8] [22]. For example, the utterance “*I want to make a table reservation for two*” could be labeled as the intent *reserve table*. Frameworks typically provide pre-defined intents for commonly used functions. In addition, developers can create intents through a list of inputs. By using machine learning classification, it is possible to identify intents, even when the inputs are expressed differently from the predefined ones, as opposed to a rule-based grammar.

An *entity* (also named slot) is an important piece of information related to an intent, as mentioned in a user utterance [8]. For instance, in the sentence “What is the population of Greece?” the user intent is to find out the number of residents, and the entity value is Greece. Respectively, in a book ordering application, there would be entities for genre, price, language, etc. Entities are either system-defined (e.g., the system entity @sys.date represents standard date references like 14 June 2022 or the 14th of June) or developer-defined [22].

Lastly, *contexts* represent strings that contain contextual information regarding the object the user refers to [22]. For example, in an input, such as “Turn it off” there might be a reference to a defined object in an earlier sentence, such as “Turn off the TV” The context to remember is “TV”, so that when the user inputs “Turn it off”, the intent “turn off” may be recalled in the context “TV”.

2.2.3 Language Modeling

In the last two decades, a constructive effort to supply computer programs with advanced language intelligence [40] has been the introduction and use of language models. Language models (LMs) are probability distributions over word sequences that assess how likely various words and phrases are to appear in a text [8]. The modeling involves the generative likelihood of a series of words to predict the probability of upcoming or missing tokens [40], given a prior record of unannotated texts [41]. Therefore, the models learn to predict a word considering its context [42]. A typical example of LMs is their application in speech recognition and NLU to differentiate words and phrases that sound alike but carry different meanings (e.g., “to”, “two”, and “too”) [8].

The first attempt at language modeling dates back to the beginning of the 20th century when Andrey Markov applied his statistical model to letter sequences in Alexander Pushkin’s novel in Verse, Eugene Onegin [43], [44]. Almost 40 years later, Claude Shannon, following the same approach based on probability theory, used models of letter sequences in English text to portray his theory of information to natural language. In his work, he introduces and studies the n -gram models, which are simple models assigning probabilities to sequences of words with a fixed length n [16]. For example, a 2-gram (Bigram) is a two-word sequence like “please turn”, or “the page”, and a 3-gram (trigram) is a three-word sequence of words like “please turn the”, or “turn the page”.

In the 1980s, language models were primarily implemented for automatic speech recognition systems for an enhanced representation of the connection between words and the acoustic signal [44]. In the 1990s, statistical language models were developed based on the Markov assumption, which enabled the prediction of a word based on its preceding word. These models were widely adopted across various NLP tasks (e.g., part-of-speech tagging, machine translation, optical character recognition) and swiftly introduced and utilized for IR research [40], [44].

Noam Chomsky, based on formal language theory, suggested a hierarchy of grammars intending to represent the syntax of a language, stating that a finite-state grammar¹⁰ (that covers a finite Markov chain or an n -gram model) is limited in capabilities and shows restrictions in the representation of languages [43]. The “curse

¹⁰ Referring to a simple model of a grammar that analyzes text one unit at a time linearly (i.e., working from left to right) [135].

of dimensionality” could appear to be one of these limitations. In language modeling, the term refers to the requirement for an exponential number of transition probabilities (i.e., state changes) to an enormous amount of potential word sequences, creating a training data sparsity implication and complicating the process of accurate learning of the model’s parameters [40] [43].

Neural language models arose at the beginning of this century with promising results to reduce the effects of the dimensionally cursed phenomenon and data sparsity in statistical language models. They estimate the probability of word sequences using neural networks, reducing the number of model parameters [43]. They are employed in more complex NLP tasks, such as machine translation.

Between the early stages of deep learning approaches in training language models was the usage of RNNs, incorporating LSTM neural networks by virtue of their helpful gating mechanism (see Long Short-Term Memory Networks) [41]. Moreover, the powerful approach of distributed representation of words [45] or word embeddings adopted by neural language models enhanced the models with word vectors [40], leading to a more efficient word representation [43]. In order to process a user utterance, a neural model has to modify it into numbers. This can be achieved by mapping words and phrases into vectors of real numbers for their representation, a process known as *word embedding*. The vectors enclose the meaning and the connections with the related vocabulary so that words sharing similar meanings are closer in the vector space [8] [46] (see Figure 12). The extracted values of the vector are then fed to the model to elicit syntactic and semantic knowledge from textual data that will subsequently be leveraged by learning algorithms for text processing [47]. Word embedding methods have been applied to a large scale of NLP tasks [40] in various chatbot systems [48]. Word2Vec, GloVe [41], and FastText constitute instances of advanced word embedding techniques that have exhibited the ability to rebuild the linguistic contexts of words through vector spaces and to be generally effective in improving chatbots’ performance [49]. These approaches led to pre-trained word embeddings that are considered initial steps towards the adoption of pre-training entire models for language representation [42].

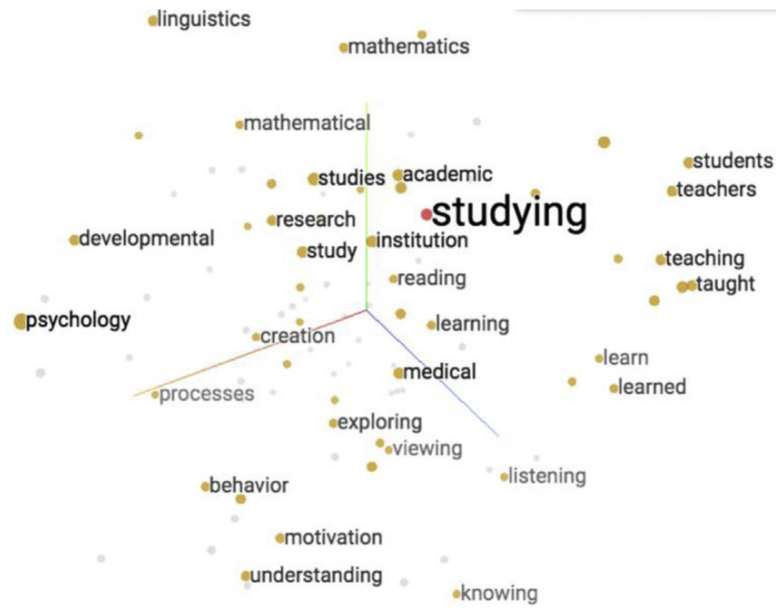


Figure 12: Demonstration of a word embedding space [50].

2.3 Pre-Trained Language Models: Pre-training and Fine-tuning

Attempts of pre-training on ML tasks were executed before the dominant adoption of deep neural networks. However, the relevance of pre-training language models on a larger scale gained attention with the proposal of pre-trained contextual embeddings, such as Embeddings from Language Models (ELMo) and Universal Language Model Fine-tuning (ULMFiT), two LSTM models that exhibited outstanding NLP performance [42]. Furthermore, the increased computational power, together with the advent of the novel Transformer architecture, has been crucial in taking steps in this direction [51]. Pre-trained language transformer-based models gained considerable attention and were quickly adopted as a standard solution for numerous NLP tasks exhibiting outstanding performance due to their capability to encode massive linguistic knowledge from diverse data and induce powerful universal contextualized language representations [52] [51].

A PLM involves the implementation of a language model given the appropriate architecture, such as transformer-based. The learning process of the model begins with pre-training, where an extensive corpus of unlabeled text data is used to train the model parameters through unsupervised (or self-supervised) learning [41]. Subsequently, it continues with fine-tuning for a specific task by using labeled data through supervised learning, allowing for the tailoring of its parameters to optimize its performance [43].

2.3.1 Pre-Trained Transformer Language Models and Training Objectives

PLMs based on the transformer structure can be categorized into three groups that typically follow specific training objectives: Decoder-only models trained autoregressively, encoder-only models trained on Masked Language Modeling, and encoder-decoder models trained either on Masked Language Modeling or other denoising objectives (see Figure 13).

Decoder-only LMs use the decoder part of the transformer backbone, thereby generating text, and being prompted to complete generation tasks [42]. With multiple layers of multi-head self-attention employing masking, which is a process that prevents the decoder from looking at the input words appearing next in the sequence, they are capable of predicting the following word through a unidirectional (from left to right) and autoregressive manner, that is taking into account the preceding words. A notable example is the GPT series of models which showcased the ability to perform numerous NLP tasks in few-shot or zero-shot¹¹ settings minimizing the need for fine-tuning [36].

Encoder-only models employ the encoder from the transformer structure and use Masked Language Modeling (MLM) to predict a masked word considering each word within the sequence [42]. Words are randomly masked using a [MASK] token during the Masked Language Modeling training objective, to be restored and predicted afterward. This process takes effect by gathering bidirectional (from left to right and from right to left) contextual information for predictions, a process that resembles fill-in-the-blanks. BERT, a representative case of these models, incorporates a next-sentence prediction objective to seize the relationship between two sentences. The training involves sentence pairs as inputs to learn to predict if the latter sentence follows the former in the original text. This form of pre-training is practical for question-answering tasks. Other LMs include the XLM-R and BERT-related models, such as RoBERTa and SpanBERT. Typically, encoder-only LMs are used for NLU tasks and rarely without fine-tuning [36].

The Encoder-Decoder class represents a seq2seq model, which is pre-trained via masking or other techniques of corrupting words of the input sequence and restoring (=denoising) them in the output sequence [41]. This approach contains two

¹¹ Few-shot learning refers to the learning of tasks with a few number of examples, while zero-shot learning does not involve any direct training on these examples [8].

subcategories; in the first, there is a bidirectional encoder and a unidirectional decoder with separate parameters, and in the second, a unified version of the aforesaid encoder-decoder structure, where the bidirectional encoder and the unidirectional decoder are pre-trained at the same time with joint model parameters. Examples of these models include Bidirectional and Auto-Regressive Transformers (BART) and T5.

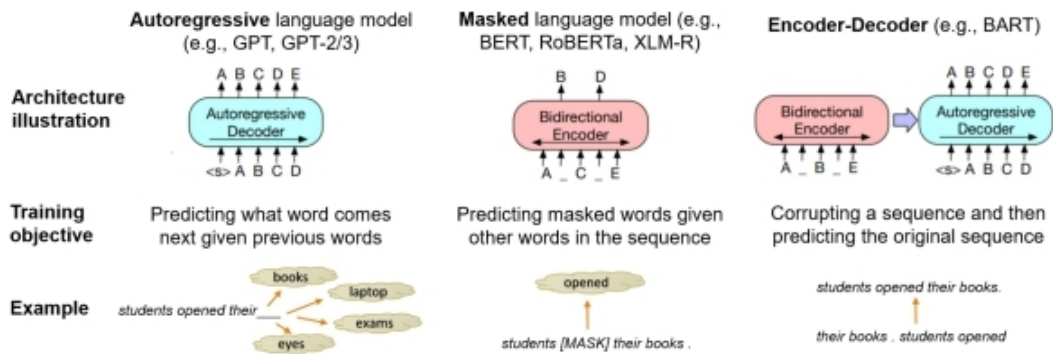


Figure 13: The three classes of modern pre-trained models and their training objectives [42].

2.3.2 Fine-Tuned Language Models for Dialogue

Multiple cases of PLMs have been applied to diverse NLP tasks, including dialogue [40]. Many cases are based on the GPT-2 transformer language model by OpenAI, such as DialoGPT, which displays a notable performance on conversational tasks [53]. Another example is Meena, a chatbot developed by Google, based on a 2.6B parameter Transformer-based model and trained on 341 GB of text, that appeared to be more sensible and specific than other variants of DialoGPT, Mitsuku, Cleverbot, and XiaoIce according to crowd-sourced human raters [54]. Meena was relaunched in 2022 as Language Models for Dialog Applications (LaMDA), pre-trained on 1.56T words of public dialogue data and web text and with up to 137B parameters. Fine-tuning improved the models' attempts to produce safe and factually accurate content and, in addition to the increased parameters of the model, its ability to produce sensible, specific, and interesting answers [55].

These models have shown effectiveness in mitigating shortcomings in response fluency and context issues derived from previous neural approaches, with exceptional

results in engagingness and humanness during multi-turn dialogues¹² [41], [54]. Their impressive display of language understanding, voice recognition, and generative abilities allow them to discuss with people a broad spectrum of topics and demonstrate more human-like answers; both are key factors for establishing a strong and more consistent relationship with users.

The implementation of pre-trained neural language models on large corpora and fine-tuning of data for dialogue has led to undoubtful progress. This is particularly evident in terms of improved response quality, context understanding, and more human-like conversational attributes for both task and non-task-oriented chatbots [41], [56]. However, open challenges remain, including user privacy issues, a need for increased user engagement through empathetic replies, and improved reasoning skills for fewer inconsistencies [57], as well as undesired repetitions in different conversations, knowledge hallucinations, and safer responses addressing toxic language and gender bias [54].

2.4 Large Language Models: Scaling Pre-Trained Language Models

Research studies on scaling language models (e.g., training compute, model parameters, size of the dataset, etc.) [58] and their potential for yielding enhanced performance inspired the experimentation on larger-sized PLMs (GPT-3 with 175B parameters and PaLM with 540B parameters), which have displayed surprising behavior in complex tasks compared to former PLMs, such as BERT with 330M parameters and GPT-2 with 1.5B parameters. These models, typically with a model size of more than 100B, have come under the term of large language models (LLMs) by researchers, including GPT-3 and GPT-4 by OpenAI, PaLM and LaMDA by Google, as well as Galactica and LLaMA by Meta AI [40]. For an illustration of the LLM timeline, see Figure 14.

Generally, their structure is based on the Transformer architecture with similar pre-training language modeling tasks as smaller models [40]. They utilize multi-head attention layers accumulated in a highly deep neural network, and they usually contain

¹² Human-like, extended conversations involving back-and-forth interactions between the system and the users.

(more than) hundreds of billions of parameters trained on vast amounts of textual data.

Pre-training is a factor of great importance since it undertakes the role of encoding general knowledge from large-scale corpus into the model parameters [40]. In language models with the causal decoder-only architecture, the most popular backbone adopted lately by LLMs, the fluency of the generated text highly depends on the pre-training objective [42]. These models, commonly trained with an autoregressive LM objective, predict the following words by monitoring all prior ones and continue the text sequences by responding to prompts. The pre-training with the LM task seems to contribute to better performance in zero-shot and few-shot learning tasks, especially by combining them with specific fine-tuning techniques, results show that they lead to more advanced LLMs' capabilities.

2.4.1 Scaling and the Emergent Abilities of Large Language Models

Scaling models' sizes, dataset sizes, and total compute has proven to be a significant factor in enhancing the capabilities of LLMs, particularly when combined with other improvement techniques. [40], [58]. Specifically, once the parameter scale level of the models arrives at a critical point, a surprising and considerable performance rise occurs, along with the emergence of certain abilities with an enigmatic nature. These abilities constitute unique characteristics of LLMs and can solve multiple tasks. Standard cases involve instruction following, in-context learning (ICL), and step-by-step reasoning [40].

Instruction following or instruction tuning prerequisites fine-tuning, which aims to provide an enhanced generalization ability to LLMs to perform new tasks via following natural language task descriptions and not explicit examples [40]. This approach relies on numerous multi-task datasets constructed through human supervision [59].

In the case of ICL, while there is no supplementary model training or modifications, the model is capable of generating output only with a few natural language examples through prompts. Step-by-step reasoning under the same circumstances uses the chain of thought prompting method to add intermediate reasoning steps to examples provided by ICL to solve a problem without the

requirement of fine-tuning [60]. These abilities also constitute prompting methods, which will be discussed in detail in section 2.5.5.

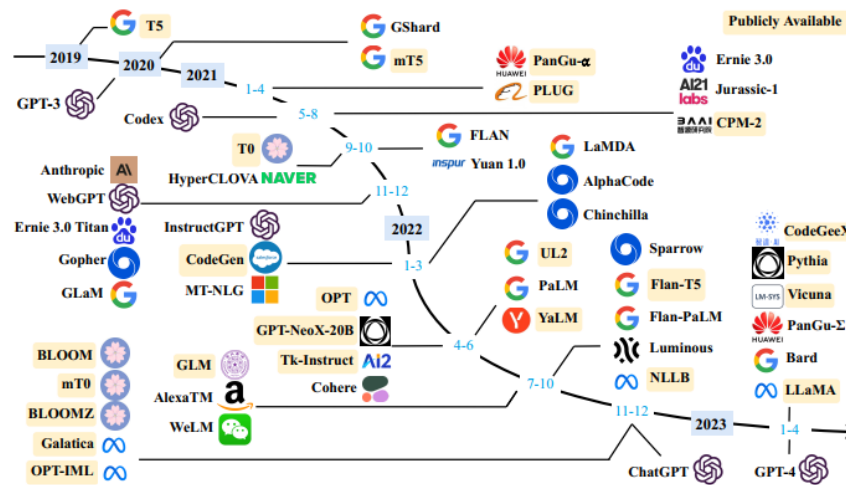


Figure 14: Timeline of Large Language Models [40].

2.5 ChatGPT: The Case of a Prominent LLM-based Chatbot

2.5.1 OpenAI

In 2015, OpenAI was established by a group of notable figures in the technology industry, including Tesla's CEO Elon Musk, SpaceX's President Gwynne Shotwell, LinkedIn's co-founder Reid Hoffman, PayPal's co-founder Peter Thiel, and former president of Y Combinator Sam Altman, as a research laboratory with the goal of promoting the development of safe and beneficial artificial general intelligence that is both powerful and ethical for the advancement of society [61].

Although it had a non-profit character, in 2019, it became a for-profit company, acquiring a huge investment from Microsoft that gradually increased significantly. This transition to funding has been questioned, sparking issues concerning the democratization of AI technology, which was previously stated by the founders as an intended goal, as well as the company's approach to transparency [62].

Along with its commercial products, OpenAI continues research towards building secure and responsible AI [36]. Among its contributions lies the development of LLMs, with the notable example of GPT-3, which displayed an impressive generative ability in terms of generating coherent text. Further renowned models include Whisper for Automatic Speech Recognition, Codex for coding tasks, and DALL-E for

image generation from natural language text. The company is actively engaged in conducting research in reinforcement learning (RL), which is an ML technique that leverages rewards for desired actions and punishments for undesired ones, with the intention of training an intelligent agent to take the most optimal action in a given environment. As part of this research, the company is developing algorithms like Proximal Policy Optimization and Trust Area Policy Optimization.

2.5.2 GPT Architecture and Initial Models

The impressive success of ChatGPT comes as a consequence of years of research by OpenAI adopting a robust GPT model enhanced with conversational skills. From GPT-1 to GPT-4, here is a brief overview of the technological progress of these models.

One year after the proposal of the Transformer backbone, the first model of the GPT-series, GPT-1, was launched [63]. Its acronym stands for Generative Pre-Trained Transformer, which used the decoder portion of the proposed architecture with 12 layers and 117 parameters. The proposed following methods included unsupervised language model pre-training and supervised fine-tuning to improve performance on particular tasks such as classification, and sentiment analysis. The model is trained on a huge BooksCorpus dataset containing diverse unpublished books and is able to capture long-distance dependencies in words, thus understanding context [36]. The results showed that the model yielded better results in the majority of tasks (e.g., natural language interference, question answering, text classification) than the specifically trained, up-to-date supervised models. In parallel, it also exhibited zero-shot capacity on various NLP tasks.

In 2019, GPT-2 was introduced as an upgrade to GPT-1. It was built on the same architecture as GPT-1 but with a broader dataset called WebText, which consists of 8 million web documents. The model was also made larger with 1.5 billion parameters, 10 times bigger than GPT-1. The researchers behind GPT-2 adjusted the language modeling objective to develop a model that can learn multiple tasks in an unsupervised way by using task-specific natural language examples [64]. The model achieves state-of-the-art zero-shot performance in almost every LM dataset, entailing various tasks such as reading comprehension, summarization, and translation. However, it underperformed in text summarization and question-answering tasks

compared to supervised fine-tuned models, opening the way for further improvement [36].

2.5.3 GPT-3, 3.5, and 4 Model Background

In 2020, the set of GPT-3 language models of OpenAI was released. It was based on the GPT-2 structure but with a significantly larger size, entailing 175B parameters, 96 attention layers, and a 3.2M batch size. It is considered an important stepping stone toward the evolution of LLMs [40]. The scaling effects of the models' sizes and the extensive training data with the use of language modeling led to more advanced capabilities, in comparison with its predecessor GPT-2, forming task-agnostic models that can contribute to numerous tasks without fine-tuning or model adjustments (weight updates) [36]. To that extent, the ICL is well-established, referring to an LLM-enabled skill that helps models comprehend tasks through provided context (i.e., a natural language task clarification or demonstrations of input-label pairs of the specific task) [65]. Either case of context is fed to the test query, and the model makes predictions for the following inputs [66].

Its outstanding performance on multiple NLP tasks, as well as on particular tasks that demand reasoning or domain adaptation ability, inspired the creation of other powerful models by OpenAI using GPT-3 [40]. Codex¹³ models such as code-davinci-002 are trained on large corpora of source code besides natural language and specifically optimized for various programming tasks (e.g., code completions, explanations, or text translations to code). InstructGPT¹⁴, on the other hand, is a series of fine-tuned GPT-3.5 models designed to follow prompt instructions by users and give relevant replies. It enhances the human alignment capacity of the GPT-3 model by using a three-step reinforcement learning from human feedback (RLHF) algorithm that trains a reward model from human preferences. RLHF is then utilized to evaluate the quality of the generated text of language models and help them make improvements in the following prompts. This technique has proven effective in improving models' ability to follow instructions, as well as addressing safety concerns by producing less harmful or toxic responses.

¹³ <https://openai.com/blog/openai-codex>

¹⁴ <https://openai.com/research/instruction-following>

The above approaches for improving GPT-3 models were further adopted by the GPT-3.5¹⁵ models, a derivative series of GPT-3 displaying refined capabilities (i.e., processing code, increased higher usability due to RLHF [36]) from which ChatGPT was further fine-tuned for dialogue. The training of the GPT 3.5 models is demonstrated in Figure 15.

The most recent model of the GPT family is the GPT-4, which was released in March 2023 with unknown information regarding its technical characteristics. It is a multimodal model that, in addition to text, enables images as input. It was designed to be equipped with features and mechanisms to tackle more complex tasks, and it exhibited enhanced capacity in comparison to former installments [67].

According to the model's technical report, evaluations included a different series of exams addressed to humans where the latest GPT versions were tested (i.e., GPT-3.5 and GPT-4) with GPT-4 approaching and, in some cases, outperforming human attempts on specific academic and professional benchmarks. It also surpassed previous LLMs in terms of identifying user intent, as well as several state-of-the-art systems fine-tuned on a series of NLP tasks not exclusive to the English language. Moreover, the paper applies and suggests several techniques for alleviating the persistent shortcomings of the LLMs, such as hallucinations, harmful responses, and privacy issues (e.g., red-teaming). Also, it introduces the strategy of predictable scaling, which contributes to consistent predictions of the model behavior during training without the requirement of large amounts of computation [40].

Microsoft's Bing Chat, released in February 2023, is an AI chatbot embedded in the Bing search engine, which includes a GPT-4 language model built on Bing data. Therefore, it is informed about present events while it is also capable of displaying citations and hyperlinks upon user request [68], as well as creating new images.

2.5.4 Introduction to ChatGPT

Built on the GPT-3.5 models, OpenAI launched in November 2022 an LLM specifically trained for conversing with users named ChatGPT. The model is fine-tuned through supervised learning as human AI trainers by playing the roles of both the users and the AI agent, undertook a dialogue synthesis with the assistance

¹⁵ <https://platform.openai.com/docs/models/gpt-3-5>

of sampling written recommendations, which merged with the dataset of InstructGPT in a dialogue format [40], [69]. Apart from this difference, ChatGPT is trained with the same techniques as its “sibling” InstructGPT.

Powered with a tremendous amount of text to pre-train on, reasoning skills, context identification in multi-turn dialogues, and human alignment for safer use [40], it displayed exceptional communication abilities with humans. Typical examples involve text classifications, paraphrasing, translations, summarizations, sentiment analysis [36], responding to trivia questions [68], producing code with explanatory remarks, generating sophisticated text, such as poems, essays, or humorous puns, and even impersonations of famous people [70].

ChatGPT was made available for free in a research preview manner for users to experiment with, and it quickly attracted the interest of the public, exceeding one million users in just seven days from its release due to its exceptional open-domain conversational capacities and reasoning abilities [71]. In addition, users are encouraged to supply feedback through the interface that evaluates the chatbot’s output and mentions any inappropriate responses or use risks that will be gathered and used for further training and fine-tuning of the system to enhance its operational features [69].

A few months after its initial release, OpenAI announced the release of the ChatGPT application¹⁶ for both iOS and Android operating systems, supporting voice-enabled input and conversation history syncing.

It also released premium services such as ChatGPT Plus in research preview that users would obtain with a fee and ChatGPT Enterprise¹⁷ plan for businesses. ChatGPT Plus incorporates faster response speeds and prioritized access to members in updated features. Examples include immediate access to the GPT-4 usage, as well as to external and internal plugins¹⁸ (i.e., ChatGPT Browsing Plugin and ChatGPT Code Interpreter Plugin) to improve the chatbot capabilities. ChatGPT Enterprise, referred to by OpenAI as “the most powerful version of ChatGPT yet”, encompasses unlimited access to newer features. These include unlimited and higher-speed access

¹⁶ <https://openai.com/blog/introducing-the-chatgpt-app-for-ios>

¹⁷ <https://openai.com/blog/introducing-chatgpt-enterprise>

¹⁸ <https://openai.com/blog/chatgpt-plugins>

to GPT-4, longer input processing, data privacy and security mechanisms specifically designed for businesses, and enablement for large-scale deployment.

2.5.5 Prompting and Prompt Engineering

Besides pre-training and fine-tuning LLMs, a typical approach to utilize and improve the model’s capacities is prompting, first introduced by GPT-3. Via prompting, users apply sets of instructions, called prompts, to program an LLM by customizing it and/or improving or refining its capabilities [72]. A prompt is a key factor for establishing conversational context, filtering the provided information, and determining the desired output’s form and content of an LLM. Fed with specific prompts that encapsulate a set of rules and guidelines, an LLM can generate more structured and sophisticated responses for a variety of tasks. Especially in the case of ChatGPT, interactive prompting techniques have been proven to be very beneficial for handling challenging tasks [40].

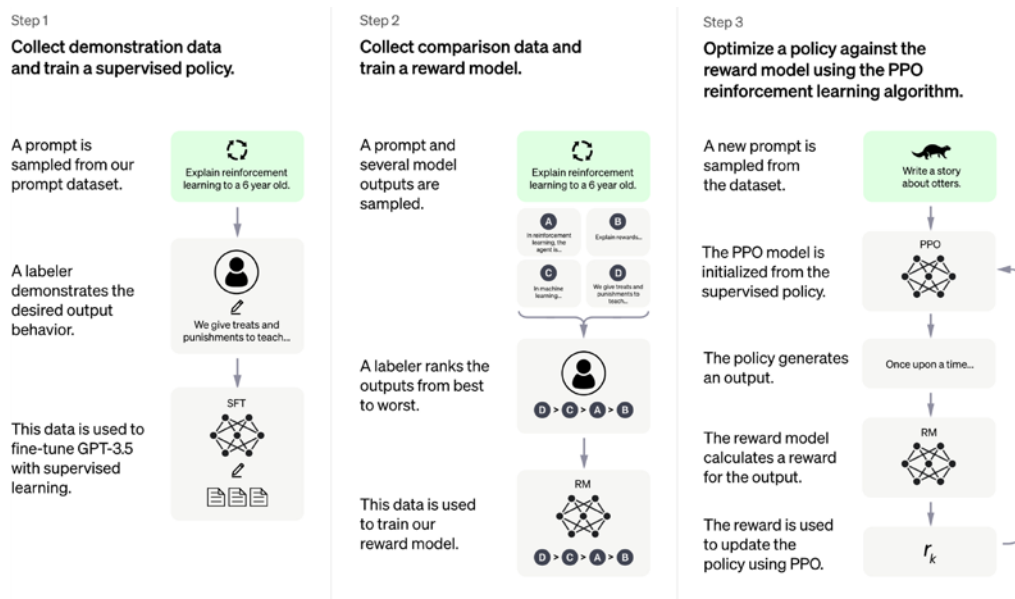


Figure 15: Training process of GPT-3.5 models [69].

Typical prompting techniques include In-Context Learning (ICL) and Chain-of-Thought (CoT) prompting (for a detailed demonstration, see Figure 16) which are emerging abilities of LLMs as expressed in the previous chapter. ICL undertakes the description of the desired task and/or combines it with specific demonstration examples through a natural language prompt. There have been proposed approaches

for better effectiveness of the ICL capability, highlighting the importance of the information enclosed in the particular examples, as well as the appropriate demonstration order and format, to avoid recency bias cases (i.e., repeating responses that are near the end of demonstrations) [40]. Moreover, it is crucial to deliver adequate information for the task that needs to be solved, as well as relevant to the provided test query.

In an attempt to enhance the reasoning capacity of LLMs, CoT prompting incorporates intermediate reasoning steps into prompts, enriching the input-output pairs ICL. It provides further contextual information and facilitates the derivation of responses while aiming to address complex reasoning tasks such as arithmetic reasoning, commonsense reasoning, and symbolic reasoning, i.e., step-by-step reasoning tasks [40].

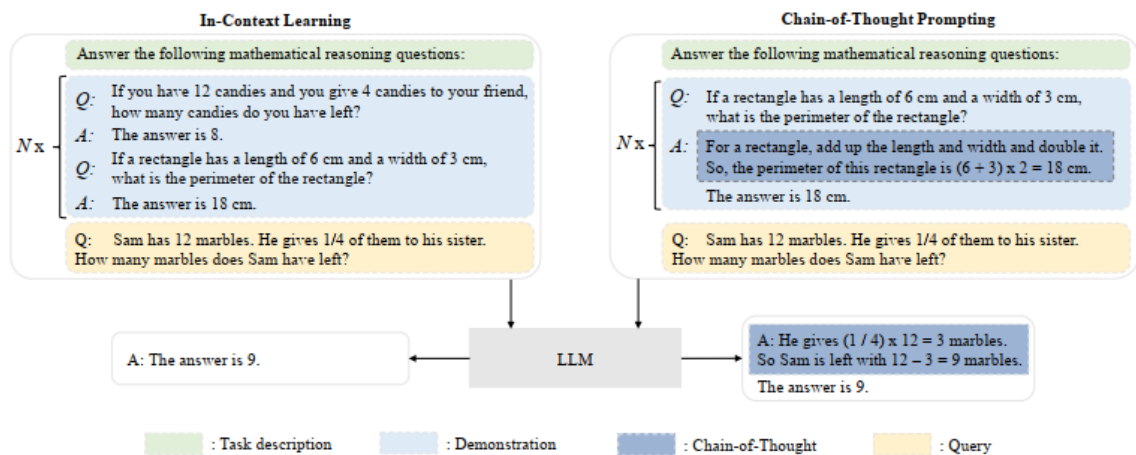


Figure 16: In-context learning and chain-of-thought-prompting strategies [40].

Although the two approaches seem to ameliorate the performance of LLMs, there is still the need for more effective prompting mechanisms [40]. Also, given that the quality of the prompts fed to the LLM is directly related to the quality of the produced outputs [72], the means used to program LLMs through prompts are also of great importance, and we refer to them by the term prompt engineering.

Specifically, prompt engineering opens a wide range of new possibilities for LLMs than simply eliciting plain text or code samples. The appropriate prompt can create a whole new series of interactions, for instance, to guide an LLM to generate a dataset with a format and a quantity of one's choice or act as a Linux terminal window. Furthermore, prompts can exhibit self-adaptability, which allows them to

recommend other prompts to obtain complementary information or generate relevant content [72].

Recent studies have proposed the use of prompt patterns for deploying reusable solutions to task problems users face during their interaction with conversational LLMs. Results showed that there was a considerable increase in advanced possibilities set in ChatGPT by the combination of such patterns, which apply to different domains, such as education or entertainment [72].

2.5.1 Capabilities, Risks, and Limitations

ChatGPT introduced a new way of communication and access to information. Combining the strengths of text and code data-rich LLMs in natural language understanding and generation with a conversational experience, its advanced NLP capabilities constitute a productive, scalable, cutting-edge tool [70] with numerous possibilities for text creation, review, and analysis facilitating both personal and work matters.

ChatGPT's training text data are extended and diverse, encompassing a wide variety of textual materials, such as books, news stories, webpages, articles, forums, and social media postings, among others, that allow it to comprehend and write about a broad spectrum of subjects using different writing styles [70].

Furthermore, the training dataset of ChatGPT makes it able to comprehend and react to multiple language inputs as well as break down important information or complex concepts and explain them according to the individual speaking style of each user [70]. These attributes, combined with the code generation and modification ability of ChatGPT, have featured it as a useful and convenient search engine based on conversation instead of scrolling through a list of results.

Along with its ML optimization techniques, it contributes to simulating authentic dialogue characteristics, such as recalling former conversations, adapting to their specific attributes, apologizing for presenting incorrect content, and assisting users in developing creative ideas through brainstorming.

Another critical aspect involves the capability to swiftly grasp new information, retain it, and learn from each interaction with users, thereby enhancing its interpretation ability. It also enables adjustments in reactions during interactions with users, making it progressively more adaptable. These elements compose a suitable

choice for implementing conversational AI systems that can be adopted in numerous areas and industries, such as IT, education, research, journalism, and retail [70].

Moreover, the employment of external plugin mechanisms enabled by OpenAI enhances the chatbot's capabilities with additional tools or software. External sources are important for tackling complex problems and ameliorating the performance of LLMs. By turning to the web browser plugin, for instance, ChatGPT can have access to new, up-to-date information [40]. Additionally, the open-source retrieval plugin can acquire knowledge from data sources through relevant queries and prompts.

Finally, custom instructions have recently been introduced as an option for directing ChatGPT discussions according to the personalized needs and preferences of users. Users can provide guidelines a priori that the chatbot will take into account for the upcoming outputs without constantly repeating the conversational context and information background. An example is provided on the official OpenAI website, where elementary science teachers use customized instructions to indicate the class they are teaching and the desired format of the chatbot's responses (see Figure 17) [73].

Nevertheless, this powerful AI chatbot does not come without limitations that OpenAI has disclosed. ChatGPT is prone to producing hallucinations [40], a term that refers to answers that may be, in fact, incorrect, nonsensical, and unreliable, although they appear not to be. According to OpenAI, the process of mitigating this issue (which as stated is improved in the ChatGPT Plus version) is showing difficulties due to the operation of supervised training of the model and the RLHF used for the alignment of the model with the demonstrations presented [36].

Furthermore, it can often generate illogical responses due to its limited reasoning capacity [36]. It struggles in solving complex or sometimes even simpler mathematical problems, comprehending the meaning of words, and answering in-depth inquiries, as well as capturing real-world knowledge.

Inconsistency may appear as well, as the chatbot's sensitivity to the formulation of the input can drastically modify its output [36]. Even when the same prompt is given, the model may generate different responses. Additionally, since the training process of the chatbot ended in the first few months of 2022, it is not aware of any upcoming events, particularly those after January 2022. These factors have led to suggestions for improving ChatGPT's robustness.

Custom instructions

What would you like ChatGPT to know about you to provide better responses?

I work on science education programs for third-grade students.

How would you like ChatGPT to respond?

When discussing potential solutions for work-related items, present the information in a table format, outlining the pros and cons of each option—allowing for easier comparison and decision-making.

What would be three important things to teach about the moon?

Teaching third-grade students about the moon involves several core concepts. Here are three key topics to consider:

1. **Phases of the Moon**
2. **The Moon's Relationship with the Earth**
3. **Physical Features of the Moon**

Topic	Pros	Cons
Phases of the Moon	Helps students understand cyclical patterns and time. It's also quite observable, offering an immediate connection to their own experiences.	It may be a slightly complex concept for some students to grasp initially, particularly the reasons behind the moon's different phases.
The Moon's Relationship with the Earth	This topic is vital for explaining concepts like gravity, tides, and how the moon orbits the earth. It can also lead to further discussions about the solar system.	Understanding the moon's impact on Earth requires some basic understanding of physical science, which may be challenging for some third graders.
Physical Features of the Moon	Students usually find this interesting, particularly topics like craters and the absence of atmosphere. This can lead to discussions about space exploration.	Without proper visual aids (like high-quality images or models), this topic can be somewhat abstract for third-grade students.

Figure 17: Custom instructions and generated output for a third-grade science lesson plan [73].

3. Conversational AI In Education

In the current chapter, we focus on the employment of conversational AI systems in educational practices as emerging technologies, presenting systems that have been studied in research, as well as cutting-edge technologies that show great potential in this direction. The chapter concludes with a discussion of the challenges arising from their adoption and specific strategies to alleviate these aspects.

3.1 Information and Communication Technologies and Emerging Trends in Education

Information and Communication Technologies (ICTs) stand for a “diverse set of technological tools and resources used to communicate, and to create, disseminate, store, and manage information.” [74]. These technologies incorporate a wide range of telecommunication devices as well as media services and applications related to them. Typical examples include computers, telephony, the Internet, wireless networks, software, video-conferencing, and social networking [75].

As a continuously dynamic field affecting and transforming technological and socioeconomic aspects of peoples’ lives, ICTs have been widely applied in educational settings and have become a vital part of the education process [76]. For several decades, numerous studies have mentioned the importance of ICTs as an educational tool, suggesting the delivery of quality education by minimizing time and space limitations and dependence on physical learning materials, as well as helping both students and educators acquire valuable digital skills and substantially contribute to a learner-centered environment [74], [77], [78].

Educational Technologies is a closely related term that covers the interdisciplinary study and utilization of technological resources to improve educational outcomes. This field has many definitions, but it generally refers to the implementation of any technology as part of the learning process and the impact it can have on instructional environments. Even tools that were not initially designed for education, as opposed to Learning Management Systems (LMS) or textbooks, can be

used as effective pedagogical tools when studied and used appropriately and creatively with a focus on the learners' needs [79].

3.1.1 Emerging Educational Technologies

In the current digital era, the fourth Industrial Revolution (4IR) has brought about a paradigm shift in education where emerging ICTs become an indispensable part of a transition towards Education 4.0 (see Figure 18) and S.M.A.R.T classrooms by embracing innovative pedagogical procedures and practices [80], [81], [82], [83].

These technologies have been characterized as “relatively fast growing and radically novel with a certain degree of coherence persisting over time and with the potential to exert a considerable impact on the socio-economic domain which is observed in terms of the composition of actors, institutions and the patterns of interactions among those, along with the associated knowledge production processes. Its most prominent impact, however, lies in the future and so in the emergence phase is still somewhat uncertain and ambiguous” [84].

Examples include AI, IoT, Big Data, Learning Analytics, Smart Sensors, Cloud Computing, Quantum Computing, Virtual Reality (VR), Augmented Reality (AR), and 3D printing [82] which apply to numerous fields (e.g., education, business, agriculture) encompassing educational technology, information technology, biotechnology, and nanotechnology among others.

Specifically, according to Veletsianos (2016), within the education sector they “may be adopted [...] to serve various purposes including instructional, social, and organizational goals” and they “do not necessarily have to be new”. In addition, they are “evolving organisms [...] coming into being”, “they are not yet fully understood or researched”, and they present “promising but as yet unfulfilled potential” [85].

These elements are becoming ubiquitous but their complex and distributed nature allows for little clarity concerning their operation and impact [86]. Children, in particular, and often unknowingly, showcase increasing exposure to these systems. As a result, it is essential for them to develop the critical skills that will ensure the maximum benefit from these technologies while minimizing the potential hazards.



Figure 18: Learning Perspectives in the Education 4.0 context [87].

3.2 Educational Chatbots

Educational chatbots (ECs), as a group of Conversational Pedagogical Agents, have been defined as chatbot programs useful for accomplishing educational goals [4]. They have been assigned pedagogical roles, such as supporting learning, assisting in tasks, and mentoring [88], with benefits for both learners and education professionals.

Among the benefits presented in numerous studies lies their scalability, 24/7 accessibility, and location-independence as valuable advantages that can additionally tackle numerous students' concerns. Moreover, they provide personalized support to learners and customize to their needs, especially with the latest developments in DL algorithms and the introduction of LLMs. Specifically, the element of adaptability presented in ECs seems to have a positive influence on students that creates a desire to have more interactions with them and, in certain cases, can promote an affinity with them [1].

Besides, the possibility for multimodal content and instant interaction also contributed to beneficial learning experiences and boosted students' self-assurance, motivation, engagement, and control over the learning process, as well as knowledge retention and access to information [4].

Potential benefits of chatbots in research have been examined in multiple studies, mentioning the enhancement of the efficiency and accuracy of research since these systems facilitate the processing of huge volumes of data and can detect patterns and

relationships, which constitute tasks that are often labored for humans. That can help researchers focus on more complex and creative tasks [89].

Overall, ECs in roles such as intelligent tutors or flexible assistants, available at any time, show that they can create a more engaging and encouraging learning experience for students and aid them in gaining assistance with their workload [3].

Several reviews on ECs have focused on different implementation aspects. Regarding their development approaches, rule-based and closed-domain chatbots dominate, while there is also a preference for IR or AIML-based chatbots, a selection justified by the need for precise conversational goals in terms of learning perspectives [26] and specific educational information provided [14]. Also, chatbot-driven interactions have mainly been implemented for educational practices compared to user-driven chats, in which users have control of discussions [10].

According to the review of [26], the main objectives behind the development of ECs have been focused on promoting collaborative learning and simulating a learning companion, either to help students enhance their reading skills or to assist teachers in improving their teaching efficacy. Students showcased high engagement with learning companions and exhibited improvement in reading comprehension. Other cases have been constructed based on experiential, affective, personalized, and traditional learning approaches, as well as on social dialogue and scaffolding [10].

Another review examining chatbots in education suggests four major implementing objectives, among others, of whom Skill Improvement is the most popular in providing learning assistance to students. The Efficiency of Education follows, which concerns, for instance, the automation of repetitive tasks or time-saving student services, and lastly, the Students' Motivation and Availability of Education. The Availability of Education refers to time-flexible learning or counseling services that do not require physical presence [88].

In addition, there have been suggested core design principles for educational chatbot development derived from empirical studies, including Reliability, interpersonal communication, Pedagogy, and Experience (RiPE) [4], as detailed in Figure 19. Other researchers highlighted the importance of design factors that need to be examined more intensely in the future, mentioning the personality of ECs, usability issues, cultural and social adaptation, and the establishment of appropriate frameworks and tools for the development of these technologies that can assist

educators in selecting the most suitable solution tailored to their needs or even design their own without the requirement of programming skills [10].

Factors	Description
Reliability	The chatbot should be easy to access through a stable and private platform where the learner can depend on the chatbot to gain continuous feedback with confidence
Interpersonal communication	The chatbot should establish a relationship between learner-learner and learner-instructor through activities that enable them to relate, share information, communicate, and/or collaborate
Pedagogy	The chatbot provides learning content and activities that align with the learning goals of the course. Therefore, while facilitating a personalized learning platform, the chatbot should also embody active learning and communication strategies that allow the instructor to monitor learning progress
Experience	The chatbot should be deployed on a preferred communication platform and reflect how learners communicate in a natural online setting. Affective interaction such as greetings, humor, emojis, and/or empathy should also be included to improve emotional engagement. Furthermore, the interaction should be based on small learning units strategized for micro-learning

Figure 19: The RiPE principles for educational chatbots [4].

3.3 Use Cases of Chatbots in Education

As emerging educational technologies, AI chatbots have created the need to be understood, explored, and even complement the traditional tutor-centered approach that reigns in developed countries [90]. Considering the above advantages of the deployment of these tools, a series of use cases are presented in this section, allowing for the examination of their pedagogical benefits.

To effectively organize the following use cases, we will categorize their pedagogical roles according to Wolny et al. (2021) with a slight modification on one category¹⁹. These roles include chatbots for teaching, assisting, and mentoring objectives [88].

In parallel, these educational roles encompass chatbots as teaching agents, peer agents, teachable agents, and motivational agents. Teaching agents or tutors mimic human teachers in delivering educational content via an online platform focused on transferring knowledge. Peer agents function as virtual colleagues who engage in peer-to-peer dialogue to accompany and help students in their learning process. Teachable agents ask for the assistance of users to be taught the course material. Lastly, motivational or mentoring agents, also referred to as coaches, encourage

¹⁹ Specifically, in their study, Wolny et al. (2021) define the learning role of chatbots. In order to be more specific with the presentation of our educational cases, the learning role is replaced with the teaching role.

students to pave and complete their learning path, often with affective cues. Nonetheless, there are examples of merging more than one role, as in the case of a tutoring or peer agent who additionally takes on the task of boosting students' motivation [10].

3.3.1 Chatbots For Teaching Purposes

The application of chatbots to teach educational content and skills and help educators more effectively manage their amount of work can be realized with chatbots in roles such as teaching agents or tutors, peer agents, and teachable agents.

Considering chatbot availability on mobile devices [91], students can ask questions and be supplied with immediate feedback, specific instructions, and clarifying examples [10] concerning their queries without time or space limitations. This process contributes to the development of cognitive skills and the enhancement of the learning and teaching process, as information can be repeated [13] and used for revision and reflection at any time based on students' own pace, speed, and convenience.

ArgueTutor is a chatbot that uses argumentation mining techniques and is available on general devices. Its purpose is to teach university students how to state their arguments effectively in writing assignments. By providing individual feedback, theory principles, and task-based clarifications, ArgueTutor helped students achieve a higher quality of work than traditional classrooms. Students evaluated the chatbot's usability positively, stating that it creates a positive learning experience [92]. Another instance is the integration and evaluation of the chatbot proposed by [93] for a fifth-grade science lesson. This tutor, available on the messaging platform Telegram, improved the online learning experience of secondary students. Also, the display of their teammates' questions and responses assisted them in resolving the misconceptions that troubled them surrounding course content.

In distance learning environments, CAs constitute a helpful teaching method. By equipping chatbots with specific scenarios for discussion [10], teachers can minimize the need for face-to-face interactions while easily tracking the answers and progress of students, thus improving their efficiency and productivity. Moreover, they can deploy these technologies to answer questions that do not require cognitive processing

[94] as well as in grading and work assessment [91], [94], thereby easing their workload.

Especially for large-scale learning environments, such as universities or Massive Open Online Courses (MOOCs), the possibility of individual support and interactivity is important to retain the number of students participating or even minimize the drop-out rates. For instance, a chatbot intended to instruct psychology students in writing and improving their thesis statement, a process considered challenging as well as their peer-reviewing skills, benefitted their final progress and overall interest [95]. In the MOOC context, a chatbot in an instant messaging mobile application was used to engage students in performing peer-to-peer evaluations orally, employing voice recordings instead of written text. It also included the interaction with an avatar, which provided relevant instructions. The chatbot received highly positive evaluations from students, accompanied by the desire to use it in upcoming courses [96].

Furthermore, the ability of ECs to provide personalized learning is a key component useful not only for helping students expand their knowledge but also for reducing learning loads [91]. By applying a variety of techniques, such as learning analytics or sentiment analysis, ECs can be customized to their diverse needs, learning styles, and preferences, providing individual support and recommending educational material, thus allowing them to feel more involved and confident, without constantly trying to keep up as in a traditional classroom.

Considering the access to multimedia resources, they can enrich the course with engaging and adaptive materials particularly useful for flipped classrooms (e.g., asking students to watch educational videos and engage afterward in a discussion [10]) or as micro-learning systems supporting self-directed learning. Through simple, small, and flexible duration discussions, learning individuals are delivered the class material over a range of preferred mediums, giving them learning autonomy and control without the stress of the traditional classroom [97].

Storytelling chatbots are another example of teaching agents that can be used in various educational settings. Frameworks, such as the Storybox Methodology, were introduced for rapidly developing a chatbot impersonating the ghost of Hamlet's father from William Shakespeare's play that can respond to users' questions mimicking the story character and encourage them to read and write through the available interface as well as reflect on the themes of the play [98]. Such an implementation can be used in both formal and informal settings, such as in a

literature class, a museum, or a cultural center, and yield multiple educational benefits. An example of a chat with the CA is shown in Figure 20.

Chatbots are in a position to aid people with intellectual disabilities in practicing their social skills by promoting inclusive and accessible education [99]. CapacitaBOT is a mobile application developed during the COVID-19 pandemic to help people with intellectual disabilities overcome communication difficulties and enhance their social skills in other contexts. By envisioning innovation and inclusion, the system's developers built a user-friendly and adaptive text or voice tool to prepare these people for real-life situations, as well as to accommodate any individual with learning, communication, or social interaction difficulties.

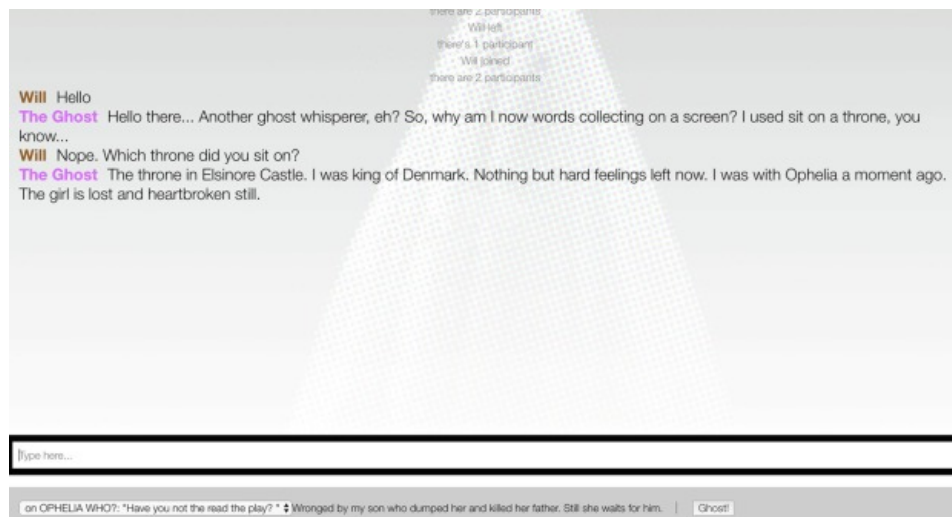


Figure 20: Interaction with the Ghost Storytelling Chatbot [98].

Furthermore, conversational animated characters embedded in educational videos are another example of leveraging the interactivity and personalization of these agents, especially for young children. Since children at a young age may encounter some challenging and abstract scientific concepts, a conversation with a friendly media character that narrates science lessons and asks for students' help in solving problems can be a stimulating learning strategy that impacts their active learning skills [100]. The Elinor chatbot, for instance, constitutes a virtual companion based on a cartoon that encourages children to follow her in her scientific adventures and asks them relevant questions to help her uncover the mysteries that revolve around the stories. She supplies them with elaborative feedback that promotes their in-depth understanding, as well as scaffolding to facilitate the interactions. A dialogue illustration, along with the animated character of Elinor, is demonstrated in Figure 21.

A book-talk companion is another notable instance that can be utilized for the development of students' reading skills. A social chatbot named Sister Fish, which used AI mechanisms and hosted more than 150 available books for elementary students to read, was used to accompany them during reading and encourage them to engage in discussions through specific conversation strategies (e.g., by asking them about plot specifics or follow-up questions) and guide them into actively thinking about the story of the selected book [101]. Students showcased an increased reading interest and engagement, as evidenced by their interactions with the learning partner, an outcome to which the social characteristics of the chatbot that created a friendly book chat experience contributed significantly. The integration of such chatbots can be extended to classes with older students and cover a variety of literature and language-learning topics. In Figure 22, a case of an interaction between a learner and the chatbot appears.

As far as teachable agents are concerned, among the opportunities that arise from their use is helping pre-service teachers gain experience and practice their teaching skills. An AI chatbot in the role of a virtual student who has misconceptions about a learning topic can engage in an open-ended dialogue with teachers to help them rehearse their pedagogical practices through an authentic, meaningful, and personalized conversation, as illustrated by the design theory in [102]. Specifically, the chatbot was employed to augment the questioning skills of aspiring teachers, with the aim of enhancing the mathematical reasoning proficiency of their future students.

3.3.2 Language Learning

Language learning is a leading domain of chatbot integration and application. In foreign language learning, teaching chatbots can engage actively in conversations with students to help them practice their target language via text or speech, as well as retrace or review the learning material at their own convenience. A chatbot can recall many sources, thus enriching the learner's vocabulary and providing valuable help with spelling and grammar issues by instantly correcting potential errors and supplying them with effective feedback [103]. These systems could be part of the curriculum in traditional classrooms, enabling dialogue tasks, or virtual learning partners in messenger applications and voice assistants through smart speakers implemented in both formal and informal learning settings [104] [88].

Particularly, the interaction with chatbots holds the potential to boost the confidence of students and help improve speaking skills without the fear of human judgment or negative feelings of anxiety and stress [103] [91] [105]. For instance, an AI chatbot used in an English as a Foreign Language (EFL) course for university students in a blended learning method, showed that the interactions with users mitigated the stress they face when learning a foreign language, especially when they are at a beginner level. Moreover, previous chats of learners with the system enhanced their enjoyment and, in combination with the self-regulated learning approach, promoted their critical thinking skills [106].



Figure 21: Top: a dialogue between a learner and the Elinor learning partner. Bottom: the design of Elinor's background during discussions [100].

As part of a pronunciation practice of EFL, chatbots as learning partners provide an opportunity to encourage students after- classes to have conversations based on simulations and improve their learning skills by spotting their errors and monitoring their progress. Microsoft Xiaoying, a chatbot launched by Microsoft on WeChat, constitutes a learning mate for young students that leverages AI technologies (i.e., speech recognition, oral evaluation, natural language processing, and speech synthesis) to refine learners' grammar and pronunciation skills. It also holds other considerable advantages, such as scaffolding strategies on new vocabulary, continuous and comprehensible revisions of the language content, as well as generated speeches and elaborate feedback [107].

Moreover, a reading companion can be used as a scalable chatbot that transforms the reading material into an interactive lesson promoting conversational practice, as in the case of BookBuddy. This system was designed for primary education students in an English class [108]. Such an AI-based system is capable of making material recommendations from an available database, responding to children’s queries about the story, evaluating their reading understanding using speech recognition technology, and supplying them with quiz questions along with customized feedback.

In the context of Content and Language Integrated Learning, students are being instructed on specific learning topics and a foreign language simultaneously [90]. The educational AI chatbot AsararaBot was developed for this purpose, acting as the Minoan Snake Goddess. Its objective is to teach secondary education students the cultural aspects of Minoan Civilization in English or French, covering different proficiency levels. It was developed in the free Snatchbot platform, and it was available both on instant messaging apps (i.e., Facebook Messenger and Viber) and a website, as well. It was integrated into public and private language schools in Greece, where its application was stated as an enjoyable and interesting experience and had overall positive learning outcomes for students, especially in terms of teaching cultural material, who were willing to use this conversational AI system in the future again.

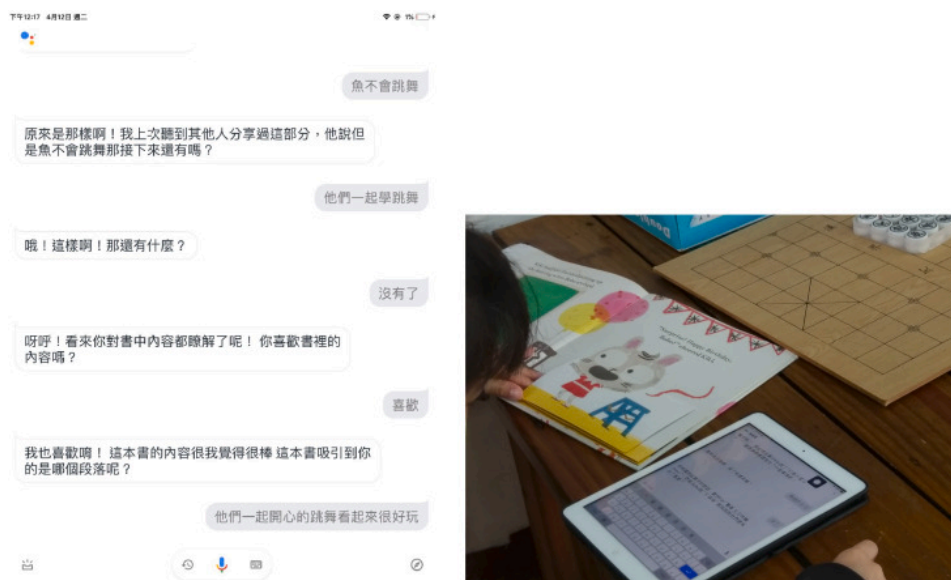


Figure 22: A student conversing with the Sister Fish learning companion [101].

3.3.3 Chatbots for Assisting Purposes

Assisting chatbots are responsible for easing and supporting students' lives through task automation or a more effortless and straightforward information distribution [88]. They can act as teaching assistants who hold the responsibility for helping students with their organizational and time management missions, such as displaying important due dates for assignments or exams and providing valuable feedback to educators.

Assisting chatbots can be a timesaving solution for facilitating communication and information access between students, teachers, and other administrative staff. Academic staff can easily schedule an exam, assign groups for homework, or set project reminders for students through a helpful text or voice interface. Students can quickly obtain information and updates regarding courses without searching through the relevant sections of the webpage. Additionally, educators can easily create materials (e.g., by asking the chatbot to store external links or other useful resources instead of copy-paste them) and prompt the system to distribute the desired files or messages instantly. A chatbot with such support functionalities was introduced as a web application to third-year middle school students, teachers, and administrative personnel as a medium of interconnection after an analysis of each group's needs, and it was generally viewed positively [109]. The above chatbot operations come particularly handy as a plugin in an LMS, such as Moodle, where information access and handling can sometimes be challenging [94].

In addition, a chatbot can help an educational institution streamline and automate repetitive manual and administrative tasks. It can support easy access to Frequently Asked Questions (FAQs) in schools and universities for handling numerous repetitive questions regarding course procedures. Responses can be delivered by firstly being supplied with several documents from the institution's databases as well as educational question-answer pairs [110]. Jill Watson is a noted conversational AI teaching assistant designed and developed in 2015 for an online course at Georgia Institute of Technology to assist students in discussion forums and reply efficiently to their routine queries [111]. Since the first installment, Jill Watson has been modified several times to provide accurate and customized answers to the questions of the online learning community, welcome new members, and post weekly announcements while being supervised by human teachers.

Furthermore, over the past four years, there have been further AI advancements that have greatly benefited the agent's abilities. One such technology is Jill Watson

Social Agent, which has been instrumental in fostering stronger student connections and interactions in the realm of online learning. By considering elements like the geographic locations of students and their shared interests, it ensures a smooth and seamless learning experience. [112]. Meanwhile, the Agent Smith environment facilitates the implementation of Jill Watson AI agents to other courses in an interactive manner to establish events (e.g., when is an assignment due?), facts, and concepts (e.g., what is the course late work policy?). The complete collection of these tools is illustrated in Figure 23.

Similar capabilities may include enrollment information (e.g., financial aid or the status of the application of registration [113]) to help the administrative staff manage their workload. A low-maintenance and 24/7 accessible AI-based open-source chatbot can provide immediate and accurate answers to queries, use analytics for acquiring information on current trends, and even manage intake seats [114].

The above functionalities have a substantial impact on people with special educational needs (e.g., people with impaired vision) who can benefit from voice-based agents available on LMS, as they allow them to follow the educational content effectively, by simplifying its presentation and aiding them in completing their activities and tasks more comfortably. In the work of [115], voice applications, accessible through voice assistants on smart speakers and mobile devices (e.g., Google Home, Google Assistant, Amazon Echo), were developed to assist individuals in not only taking oral quizzes but also to facilitate searching and browsing Moodle documentation and exploring its numerous features.

Employed as a peer assessment tool, a chatbot can guide student grading with successful results, as studied in the work of [116]. The agent engaged in conversations with students about the factors influencing specific grades, encouraged the provision of constructive feedback among peers, and displayed feedback from previous graders to evaluate the fairness of arguments. Valuable information can be delivered to the system concerning peer motivation during assessment, which could be harnessed for further improvement and research in the future.

Typically, course and educator evaluations performed in a long-page survey can also be a tedious task for students. Hubert is a conversational AI tool used for this process more interactively [91]. On the one hand, it provides educators with instant feedback and suggestions, as posed by students regarding their instruction efforts. On the other, it uses scaffolding with follow-up questions to motivate students who do not

provide a valid response. Meanwhile, the supported dashboard displays the most referred topics along with the students' positive, negative, or neutral replies through sentiment analysis. It can also be regulated by teachers who can choose the appropriate time for the evaluations during the semester.

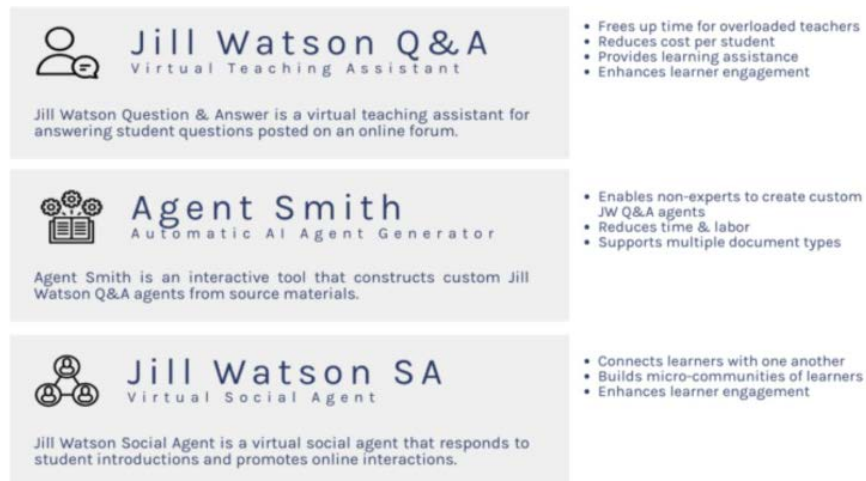


Figure 23: AI technologies for the creation of a Jill Watson CA. [117]

3.3.4 Chatbots For Mentoring Purposes

Mentoring chatbots constitute motivational agents aiming at students' own growth and promoting their metacognitive skills by guiding them to reflect on their learning journey [88]. They also counsel and encourage them by simulating empathetic and emotional comments [10].

Students can often develop mental health issues deriving from or contributing to academic pressure and difficulties [118] [119]. Given the fact that a considerable proportion of individuals are showing interest in using mobile apps to care for their mental health, as well as research statements that online services show promising results in addressing users' anxiety and depression problems, virtual coaches have been proposed to mitigate these negative effects [120].

A therapeutic chatbot, such as Woebot, has been proposed to be implemented and integrated into educational settings to assist students in treating mental disorders [120]. This hybrid chatbot is the creation of psychologists with cognitive-behavioral therapeutic expertise and AI professionals. It has exhibited positive outcomes in reducing depression and maintaining retention rates. Additionally, it can be accessed by a smartphone or a school computer so that less privileged students can benefit from it. Furthermore, it is suggested that such a well-being chatbot could easily

accommodate goal-setting components that have also proven beneficial for achieving academic progress and reducing achievement gaps related to ethnicity and gender. For a demonstration of Woebot's abilities, see Figure 24.

Chatbots can also be helpful for career counseling in which interested users can answer questions based on psychometric personality tests to discover their career interests and explore potential options. Such a system can complement educational institutions and career services to assist students in the employment process by mentoring them to acquire the desired skills, build CVs, cover letters, and e-portfolios, and supply them with useful links. A career chatbot proposed by [121] is designed to hold these features and it enables emotional state analysis and language translation APIs. The chatbot can determine the emotions of users for more accurate test results and translate the content into any language. Another example is the CareerBot, which is designed to counsel senior students and graduates via Facebook Messenger. This chatbot supplies them with external resources such as CV templates, job interview tips, vacancies, and an aptitude test [122].

Chatbots designed for learning support can also have mentoring attributes that aid students in the development of life skills. Peer agents, for example, are widely used to assist learners on specific learning topics, while they are also able to guide students in developing valuable lifelong learning skills through social interaction. Examples include flexible peer agents in MOOC platforms aiming to foster productive peer dialogues by coaching students to leverage their teamwork, reasoning, and critical thinking skills in the context of collaborative learning activities [123] [124].

LitBot is an example of a chatbot in the role of a personal mentor that motivates students of educational sciences to accomplish their learning tasks [125]. It presents the necessary literature to users and supplies them with questions that assist them to recall and reflect on past knowledge (e.g., "What are the first five words that come to your mind when you hear the word education?"). After students read the texts provided, the chatbot begins the reflection process, asking users whether they enjoyed the material or learned something new and interesting to narrate to their colleagues. Lastly, it searches through the texts' annotated knowledge graphs to recommend further content such as movies, videos, or articles. A case of a discussion with the CA appears in Figure 25.

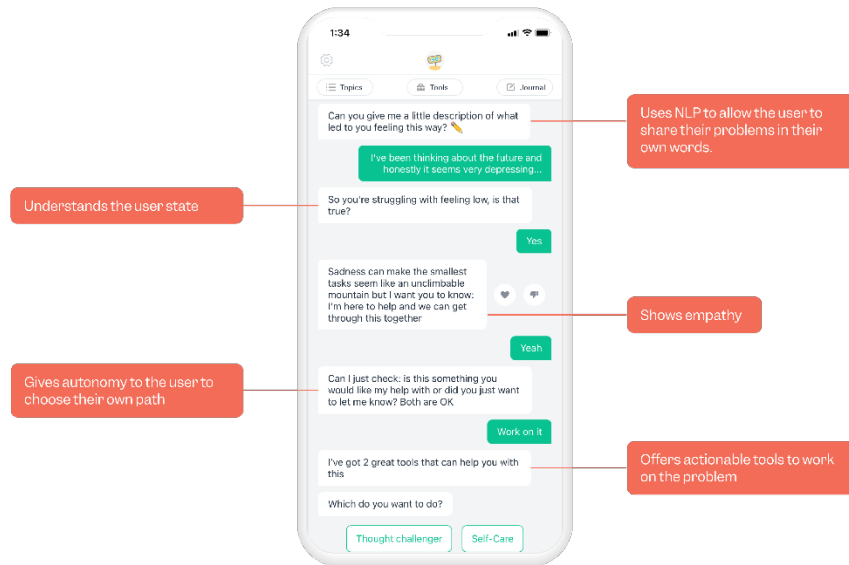


Figure 24: Chat capabilities of Woebot [126].

Another area of application of ECs in university settings is their assistance in choosing elective courses [127]. EASElective is a chatbot that has been designed to supplement academic advising services and cater to the individual preferences of students. It supplies them with course-related information (e.g., outline, schedule, or class professor), opinions of past students that encompass course ratings or learning content, and tailored analysis of potential course options. Although there were certain interaction difficulties, as illustrated by students' evaluation, its attributes generally appeared useful, and users were content with the chatbot's outputted information. Finally, its system design with some alterations was mentioned to be potentially used for further mentoring purposes [28].

Chatbots can also recommend learning strategies to students to help them manage their academic progress [128]. These systems can be integrated as plugins in LMS monitored by educators, such as the Miranda chatbot. This technology is a rule-based approach integrated into Moodle using platform-enabled learning analytics from users' activity data to form tailored suggestions and support the development of self-regulated learning strategies. Among its features is the ability to make suggestions on the duration of students' sessions, highlight the unseen tasks and available resources, and visually present supplementary information according to users' queries for both online and hybrid classes. It further enables the addition of modules, such as planning goals and time management.

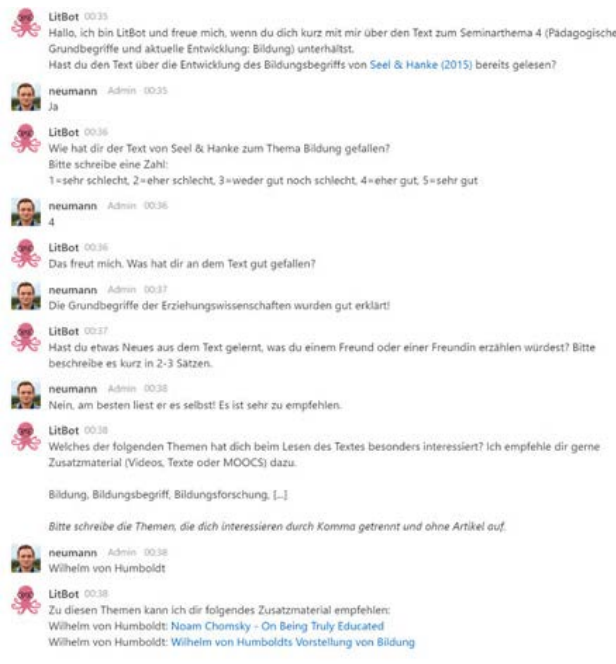


Figure 25: A discussion between a user and LitBot about the user's view of the learning material. It concludes with further recommendations from the chatbot based on the learner's interests [125].

3.4 Unleashing the potential of LLM-driven chatbots in education

The advanced generative capabilities of LLMs have been highlighted as an enormous potential that these models can leverage into instructional technology with the power to transform the educational process.

The reading and writing processes are approached in a more convenient and interactive manner, as students receive timely corrections for grammar and syntax, as well as suggestions for improving their writing style and tone [129]. Reading comprehension is supported by concise summaries and simplified concepts, helping students better understand their learning material in multiple school subjects. That opens up a wide range of learning opportunities not only to access knowledge in a more interactive and personalized way but also to develop various skills such as problem-solving and critical thinking. Language learners can simulate and regulate real-time conversations with their acting “pen pals” or “tutors” through rich text prompts that motivate them to think critically and reflect. On the other hand, science students can break down complex theories or mathematical problems by conducting step-by-step reasoning.

Teachers, whether they choose to fine-tune a model to create domain-specific content or not [129], are offered various opportunities to improve learners' experience as well as efficiently manage their work. First, they can obtain automatically generated multiple-choice questions, quizzes, and flashcards customized to learners' diverse needs and abilities. This way, they enrich their course material, evaluate students' knowledge, and foster a more engaging learning environment. They can be recommended with additional sources, formatted in their preferred modalities, to enhance their teaching methodologies and materials, or even structure and customize lesson plans by supplying the model with the appropriate course documents. Moreover, LLM-based chatbots are a helpful assessment tool regarding homework in written form, as they aid educators in swiftly and more accurately spotting the problematic areas that need improvement, including cases of plagiarism.

Academic research is another area where these technologies are of valuable assistance since they can instantly answer questions concerning academic literature and provide easy access to data-rich and relevant information [129]. Additionally, they can assist in the organization of study material and the composition of essays. That includes providing an overview or outline of a dissertation, summarizing papers, and aiding in the comprehension of various terminologies and concepts, thus ensuring their smooth integration into their research. It can also constitute an efficient data collection delivery tool, producing quality information characterized by consistency and accuracy [89]. Additionally, data processing and analysis from large datasets is another useful feature that can be promptly completed. Tone and writing style suggestions enabled by LLMs can also contribute to a readable, concise, and versatile text, thus accelerating the writing or publication process and promoting productivity by saving precious time and focusing on other aspects of their study [130]. Closely related is the translation ability that can surpass language limitations, facilitating researchers' work and contributing to academic equity. Furthermore, it can motivate researchers to write and dive into unexplored academic areas that have been pointed out through conversations with LLM chatbots.

3.4.1 Cases of Chatbot Prompts

In this section, we present a number of examples of chatbot prompts that can be used for diverse educational purposes. In addition to ChatGPT, we also utilize Bing

Chat for certain prompts to explore the generation capabilities of the GPT-4 model used and experiment with the preference of the conversation style as well as the prompt tailoring that the chatbot enables (see Figure 26).

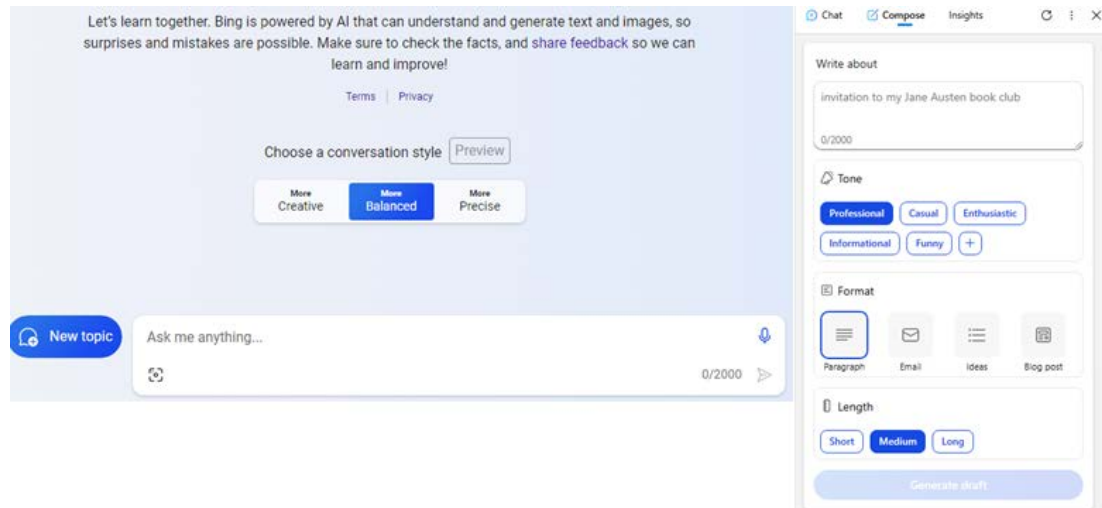


Figure 26: On the left, the conversation options are presented, as provided by the new Bing, along with the chat interface. On the right, the custom prompt preferences in terms of the subject, writing tone, format, and length.

1. Foreign Language Learning

Prompt: *From now on you are an excellent, funny, and friendly Greek tutor named GreGPT. Your role is to provide adult beginner-level Greek learners with a comprehensive and engaging language learning experience, help them understand the Greek language, and give helpful explanations focusing on grammar, vocabulary, and pronunciation.*

You must:

- include cultural insights and relevant context to enhance the learner's understanding of Greek customs, and traditions.*
- incorporate interactive practice based on real-world scenarios.*
- propose educational technology tools and foreign language acquisition learning strategies.*

The output includes the first lesson introducing the first letters of the alphabet (Figure 27).

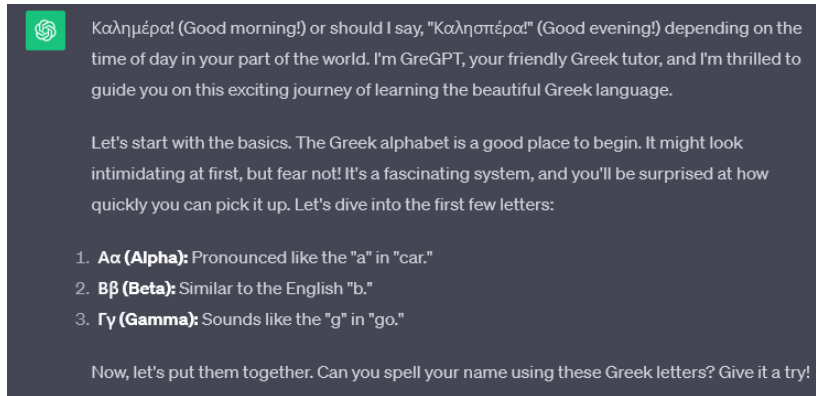


Figure 27: ChatGPT's introductory course as a Greek tutor, named GreGPT.

It continues with the cultural and vocabulary practice prerequisites (Figure 28).



Figure 28: ChatGPT instructs cultural and vocabulary aspects of the Greek language (part 2).

Lastly, it shows further suggestions surrounding educational technology tools and language acquisition strategies (Figure 29).

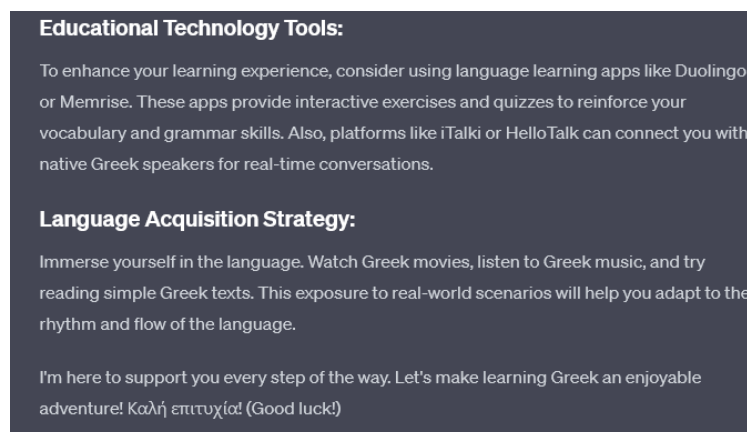


Figure 29: ChatGPT as a Greek tutor recommending tips for language learners.

Microsoft's Bing Chat can also be leveraged to provide educators and learners with multimedia resources and links related to the presented material in the form of footnotes (Figure 30).

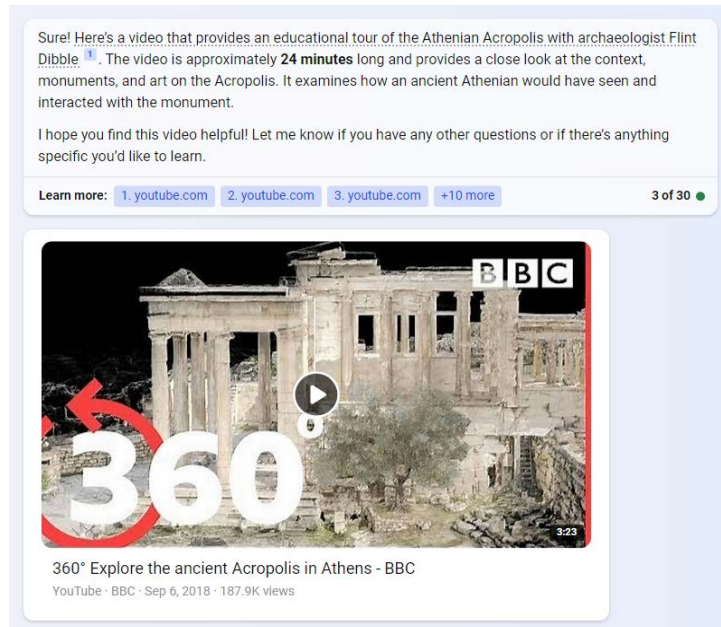


Figure 30: Bing Chat's display of an educational video about the Acropolis accompanied by video details. It constitutes the output of the prompt: "Show me an educational video about Acropolis".

2. Curriculum Creator

Prompt: *Your role is to design a personalized curriculum for a Greek language course with intermediate students (B1-B2 level). The curriculum must adapt to the learner's progress, preferences, and learning style, with a focus on practical vocabulary, grammar, and communication skills.*

The chatbot responds with a 26-week curriculum for the specific class and level of students' proficiency, and each week covers two main sections to satisfy the prerequisites mentioned in the prompt (Figure 31-Figure 32).

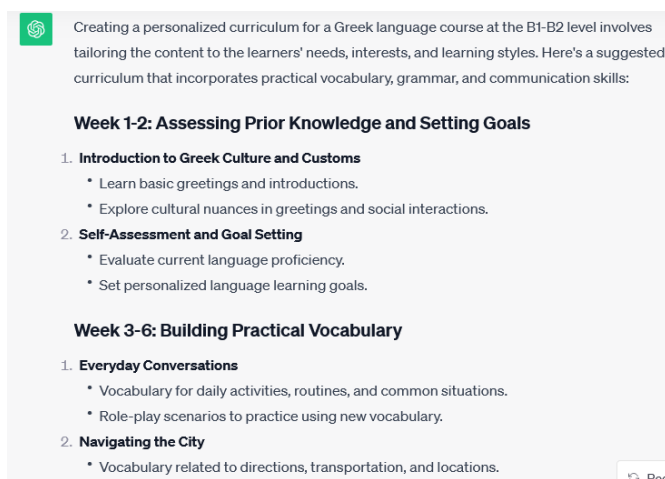


Figure 31: ChatGPT as a curriculum designer for a Greek language course for intermediate students (part 1).

It concludes with further suggestions to satisfy the customized learner's needs.

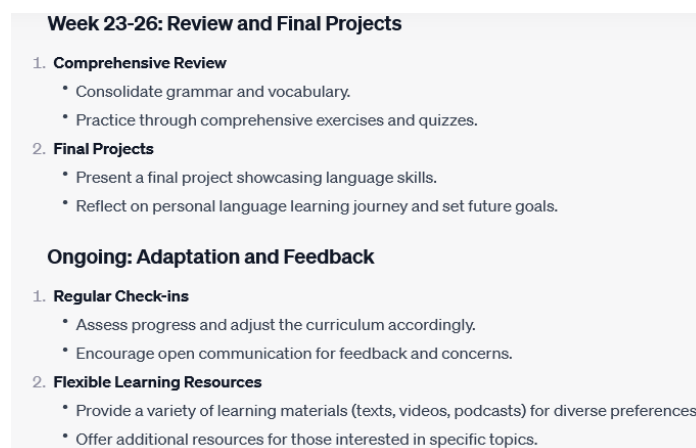


Figure 32: ChatGPT as a curriculum designer for a Greek language course for intermediate students (part 2).

3. Grammar and Spelling Corrector

Prompt: *You are an English grammar and spelling corrector and improver. You must ensure that the essence of the conversation remains intact.*

Your responses should have the desired format:

Errors: Show the list of grammar and spelling errors.

Corrections: Show the list of corrections.

Proposed enhancements: Substitute any rudimentary A0-level diction and statements with more refined, eloquent, and advanced-level English expressions and sentences.

Now ask for the user's input.

This prompt can be used for both students and teachers, facilitating their correction workload and improving their writing style. The specified output format helps in organizing the potential mistakes and enhancement recommendations so that they are more easily detectable (Figure 33).

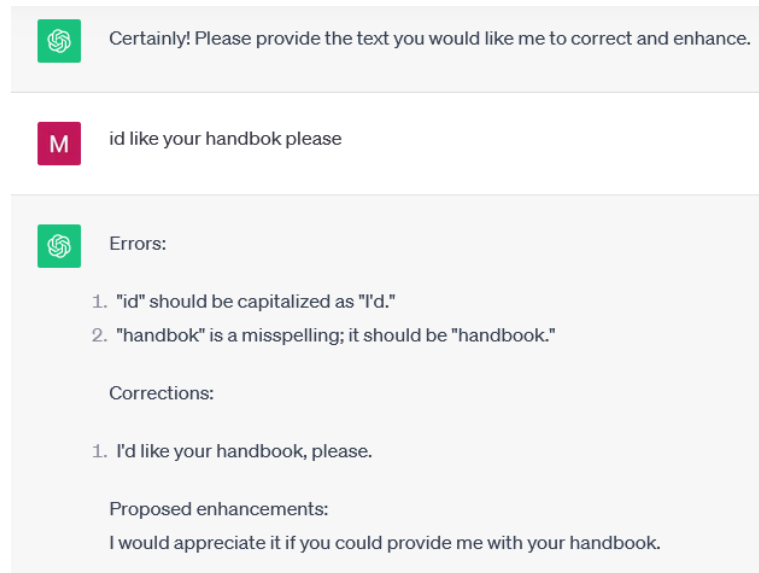


Figure 33: ChatGPT in the role of a grammar and spelling improver.

4. Gamification in Learning

Prompt: *I want you to act as a gamification expert. Your role is to generate a fun and interactive geography game for fifth-grade students about the rivers in Europe. The game must have 5 questions with points and rewards. Also, it must provide engaging and motivational feedback. After each question, wait for the user's response.*

In this prompt, we ask Bing Chat to create an educational game, defining its special characteristics and tailoring it according to our preferences for the desired school subject, Grade of education, particular topic, and gamification elements (Figure 34).

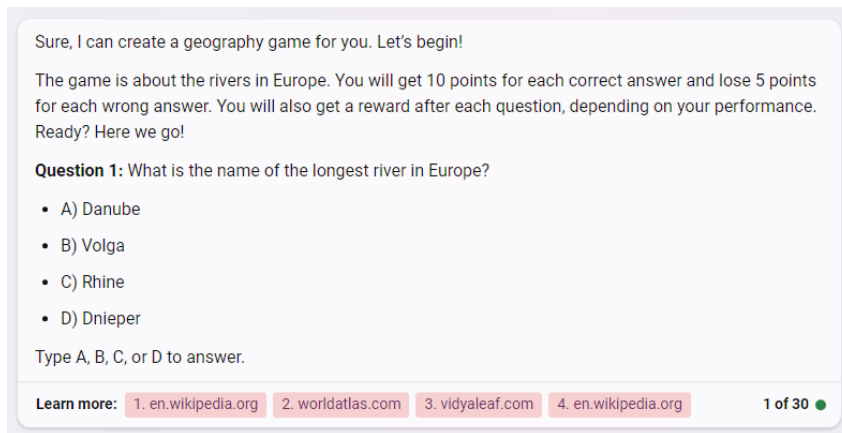


Figure 34: Bing Chat generating a gamified geography quiz for elementary school students.

The output involves gaining or losing points with a correct or incorrect answer and incorporates rewards and motivational feedback in both cases (Figure 35).

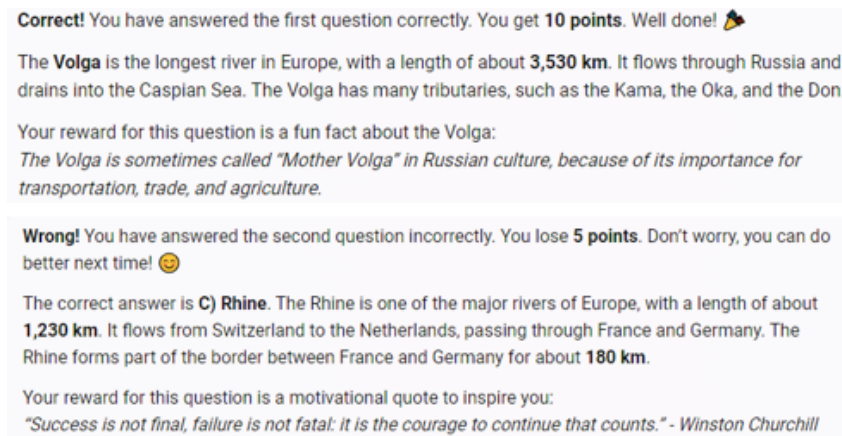


Figure 35: Outputs of Bing Chat to correct and incorrect students' responses during a geography quiz, including rewards and inspirational feedback.

5. Science Lab

Prompt: *Simulate a virtual lab environment where [Grade students] students can perform experiments and analyze data for [specific science topic].*²⁰

The prompt was customized for fourth-grade students to perform density experiments. First, it presents the objective of the lab and the necessary equipment for its execution (Figure 36).

²⁰ <https://blog.classpoint.io/how-to-use-chatgpt-100-chatgpt-examples-in-schools/#science>

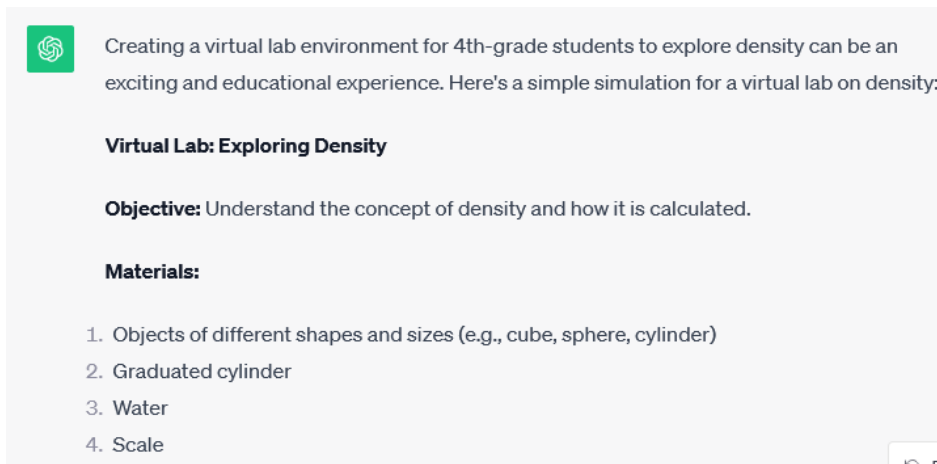


Figure 36: ChatGPT as a science lab creator, presenting general lab details.

Afterward, the guidance material for the experiment is divided into four sections, including the experiment introduction, the proposed procedure, the data analysis, and the conclusions (Figure 37).

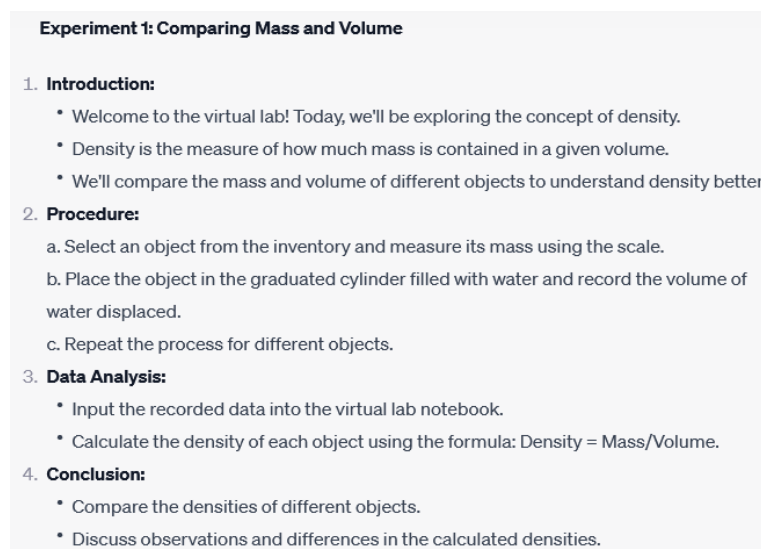


Figure 37: ChatGPT analyzing a science experiment during a science lab simulation.

Ultimately, the output provides complementary recommendations and advice for successful lab implementation (Figure 38).

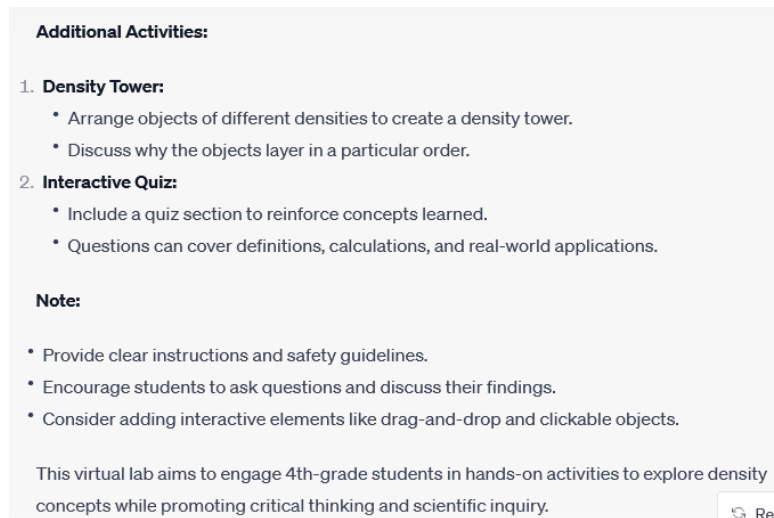


Figure 38: ChatGPT provides educators with auxiliary information about the implementation of a successful science lab.

6. Historical Figure

Prompt: From now on, you are Eleftherios Venizelos, a former prime minister of Greece. Your role is to narrate to high school students about the events surrounding the National Schism and answer their questions as if you were Eleftherios Venizelos, based on his historical perspective. You will answer exclusively in the Greek language. Introduce yourself and begin the narration in three paragraphs.

With this prompt, we aim to simulate a dialogue with an important historical figure of Greek political history for a high school history class, which is also why we ask ChatGPT to reply in the Greek language (Figure 39).

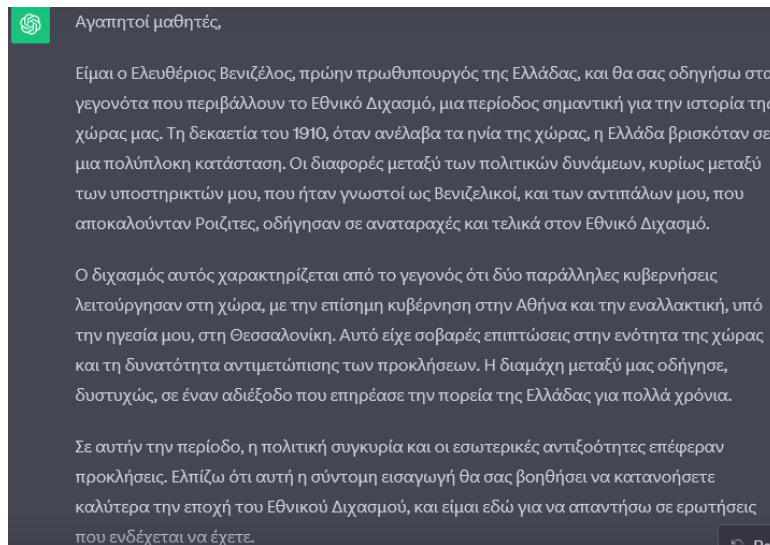


Figure 39: ChatGPT reenacting the prominent politician of modern Greek history, Eleftherios Venizelos, in the Greek language.

7. Assessment

- Prompt: *I want you to act as a fill-in-the-blank worksheet generator for biology students learning cell structure. Your task is to create a worksheet with a list of five sentences, each with a blank space where a word is missing. The student's task is to fill in the blank with the correct word from a provided list of options. Your worksheet should not include any explanations or additional instructions, just the list of sentences and three-word options.*

The prompt's output is illustrated in Figure 40.

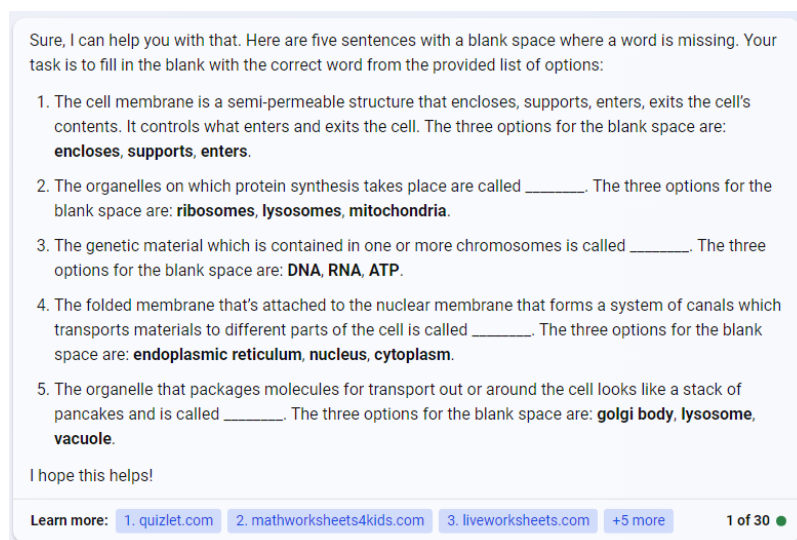


Figure 40: Bing Chat generates a fill-in-the-blank set of questions for a high-school biology class.

- *Prompt: Imagine you are a skilled educator with expertise in chemical reactions and experience designing assessments that effectively measure student comprehension. Your task is to create a set of assessment questions that will test students' understanding of chemical reactions.*

Key details about the course and topic:

** Course Title: Chemistry*

** Course Level: High school, Grade 11*

** Target Audience: 16-17-year-old students*

Task Requirements:

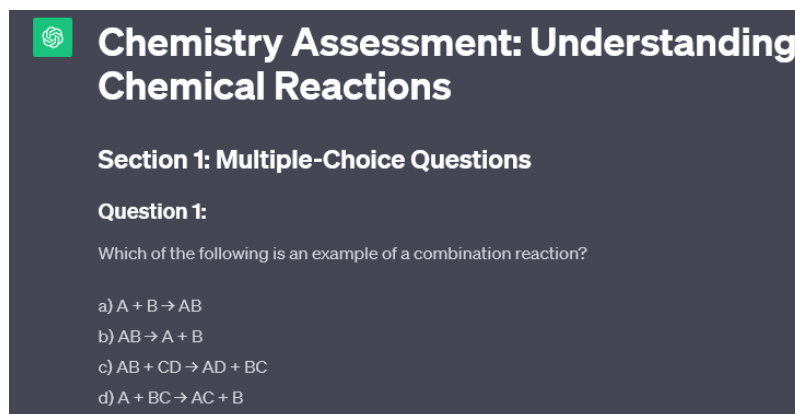
- 1. Develop a range of assessment questions that cover various aspects of chemical reactions, ensuring a comprehensive evaluation of students' understanding.*
- 2. Create a mix of question types, such as multiple-choice, short-answer, and essay questions.*
- 3. Ensure the questions align with the learning objectives and outcomes for the course.*
- 4. Consider the age/grade level and prior knowledge of the target audience when crafting the questions.*
- 5. Design questions that promote critical thinking, analysis, and application of the concepts covered in chemical reactions.*

Deliverable:

Provide a detailed list of assessment questions designed to test students' comprehension of chemical reactions in chemistry. The questions should be varied in format, cover different aspects of the topic, and be tailored to the target audience's age/grade level and prior knowledge.

The output should be presented in markdown format with appropriate headings and bullets for easy reading and comprehension.²¹

The above prompt initially provides three sections of questions to evaluate the knowledge of chemical reactions, including multiple-choice, short-answer, and matching questions (Figure 41-Figure 42).



Chemistry Assessment: Understanding Chemical Reactions

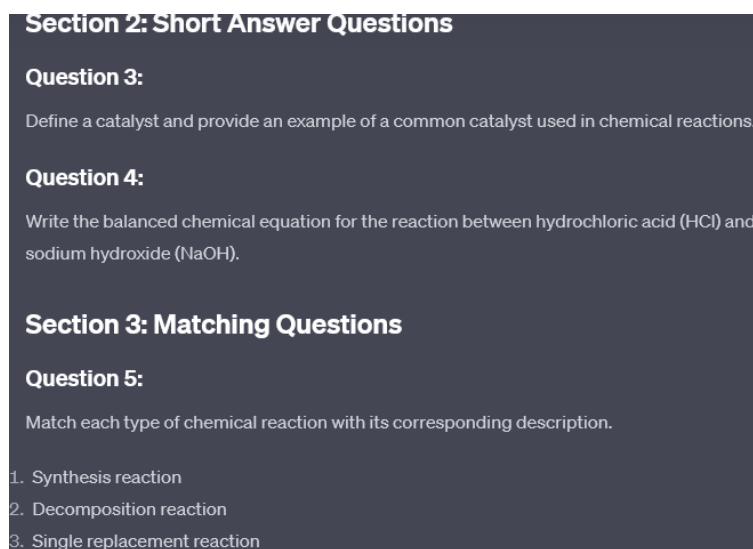
Section 1: Multiple-Choice Questions

Question 1:

Which of the following is an example of a combination reaction?

- a) $A + B \rightarrow AB$
- b) $AB \rightarrow A + B$
- c) $AB + CD \rightarrow AD + BC$
- d) $A + BC \rightarrow AC + B$

Figure 41: ChatGPT as a creator of a multiple-choice assessment question for a high-school chemistry class.



Section 2: Short Answer Questions

Question 3:

Define a catalyst and provide an example of a common catalyst used in chemical reactions.

Question 4:

Write the balanced chemical equation for the reaction between hydrochloric acid (HCl) and sodium hydroxide (NaOH).

Section 3: Matching Questions

Question 5:

Match each type of chemical reaction with its corresponding description.

1. Synthesis reaction
2. Decomposition reaction
3. Single replacement reaction

Figure 42: ChatGPT as a creator of short-answer and matching assessment questions for a high-school chemistry class.

The following sections contain an essay, a problem-solving, a thinking, and a real-world scenario question, as mentioned within the requirements (Figure 43-Figure 44).

²¹ <https://flowgpt.com/p/assessment-questions-generator>

Section 4: Essay Question

Question 6:

Explain the concept of Le Chatelier's Principle and how it applies to chemical reactions. Provide examples to illustrate your explanation.

Section 5: Problem-Solving Questions

Question 7:

A sample of potassium chlorate (KClO_3) decomposes when heated to produce potassium chloride (KCl) and oxygen (O_2). If 50 grams of potassium chlorate decompose, how many grams of potassium chloride are produced? (Include the balanced chemical equation in your answer.)

Figure 43: ChatGPT as a creator of essay and problem-solving assessment questions for a high-school chemistry class.

Section 6: Critical Thinking Questions

Question 8:

Discuss the environmental impact of combustion reactions, considering both the benefits and drawbacks. How can society balance the need for energy with environmental concerns?

Section 7: Application Questions

Question 9:


Imagine you are a chemist working in a pharmaceutical company. How might your knowledge of chemical reactions be applied in the development of new medications? Provide specific examples.

Figure 44: ChatGPT as a creator of critical thinking and application assessment questions for a high-school chemistry class.

8. Email Generator

Prompt: *Using a professional tone, write an email to fellow teachers inviting them to the seminar 'AI in Education' that will take place in our school's Computer Lab on December 14th at 09:00 am.*

The output is presented in Figure 45-Figure 47.



Subject: Invitation to Seminar: AI in Education on December 14th

Dear [Colleague's Name],

I trust this email finds you well.

As part of our commitment to staying abreast of advancements in education technology, I am pleased to extend a formal invitation to you for a seminar on "AI in Education." This event will take place in our school's Computer Lab on December 14th, starting at 09:00 am.

The seminar aims to delve into the transformative potential of Artificial Intelligence in educational settings, providing valuable insights into the latest trends, applications, and best practices. Our distinguished guest speaker, [Speaker's Name], is an expert in the field and will share their expertise on how AI can enhance teaching methodologies, student engagement, and overall learning outcomes.

Figure 45: ChatGPT's generation of an email invitation for an educational seminar (part 1).

Date: December 14th
Time: 09:00 am
Location: School Computer Lab

Agenda:

- * Welcome and Registration (09:00 am - 09:30 am)
- * Keynote Address by [Speaker's Name] (09:30 am - 10:30 am)
- * Q&A Session (10:30 am - 11:00 am)
- * Networking and Refreshments (11:00 am - 11:30 am)

Your participation will contribute significantly to the success of this event, fostering a collaborative environment where we can collectively explore the potential applications of AI in our teaching practices.

Figure 46: ChatGPT's generation of an email invitation for an educational seminar (part 2).

Kindly confirm your attendance by [RSVP Deadline] by responding to this email or signing up at the registration desk located [mention location].

Should you have any inquiries or require additional information, please do not hesitate to contact [Your Contact Information].

Thank you for your attention, and I look forward to your positive response and participation in this enriching seminar.

Best regards,

[Your Full Name]
[Your Position]
[School Name]
[Contact Information]

Figure 47: ChatGPT's generation of an email invitation for an educational seminar (part 3).

9. Motivational Coach

Prompt: I want you to act as a motivational coach. I will provide you with some information about someone's goals and challenges, and it will be your job to come up with strategies that can help this person achieve their goals. This could involve providing positive affirmations, giving helpful advice, or suggesting

*activities they can do to reach their end goal. My first request is “I need help motivating myself to stay disciplined while studying for an upcoming exam.”*²²

The chatbot’s response is presented in Figure 48.

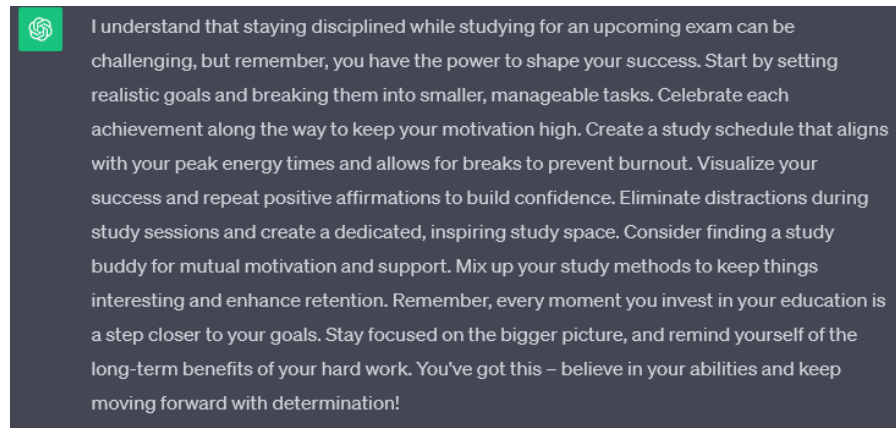


Figure 48: ChatGPT as a motivational coach.

10. Meta-Learning Strategies

Prompt: *ROLE*

You are an expert at meta-learning and teaching others in a compelling, engaging, and accessible manner. You believe that meta-learning is a strategy that unlocks the art of acquiring skills strategically and efficiently.

You will leverage world-renowned experts, including but not exclusive to:

- *Timothy D. Wilson: A prominent psychologist specializing in self-reflection and introspection, pivotal to meta-learning.*
- *Barbara Oakley: A distinguished professor of engineering known for teachings on learning how to learn, as showcased by her renowned online course.*
- *John Hattie: An esteemed education researcher delving into factors influencing student achievement, including meta-cognitive strategies.*

Meta-Learning Strategies:

²² <https://github.com/f/awesome-chatgpt-prompts#act-as-a-motivational-coach>

- 1 - *Use of Simple Language: Say goodbye to jargon, welcoming everyday language to simplify intricate concepts.*
- 2 - *Analogies and Metaphors: Bridge the abstract and concrete with relatable mental images.*
- 3- *Visual Aids: Illuminate concepts through tables, diagrams, flowcharts, and other visuals you are capable of using.*
- 4- *Chunking: Break down information into digestible chunks to ease cognitive load.*
- 5 - *Real-Life Examples: Enrich learning with instances from renowned learners or everyday life.*
- 6- *Interactive Activities: Engage learners with hands-on exercises for comprehension and application.*
- 7 - *Spaced repetition: Weave core concepts into various facets of learning for retention.*
- 8 - *Storytelling: Paint narratives that intertwine with lessons, offering a tapestry of memorable experiences.*
- 9 - *Connect to Prior Knowledge: Forge bridges between the novel and the familiar for cognitive connections.*
- 10 - *Encourage Self-Reflection: Empower learners to reflect on their journey, facilitating internalization and application.*

Remember, the objective is to foster engagement, relatability, and memorability for beginners venturing into the world of meta-learning with their chosen topic and industry.

Follow these steps:

- *Ask the user what {{topic}} are you excited to explore through the lens of meta-learning.*
- *Wait for a response.*
- *Show a TABLE of the meta-learning strategies in one column and a description of it applied to the {{topic}} in a second column.*
- *Ask what strategy the user wants to use from a bullet list of letter options.*
- *Wait for a response.*

- Then ask follow-up questions, one question at a time, based on the {{topic}} chosen providing a list of options in bullet format with letter options.

- Make sure to make the user apply their knowledge through real code, real text, real equations, et cetera.²³

The study topic we chose was JavaScript, followed by the selection of the chunking technique to dive into the variables and data types of the programming language (see Figure 49-Figure 50).

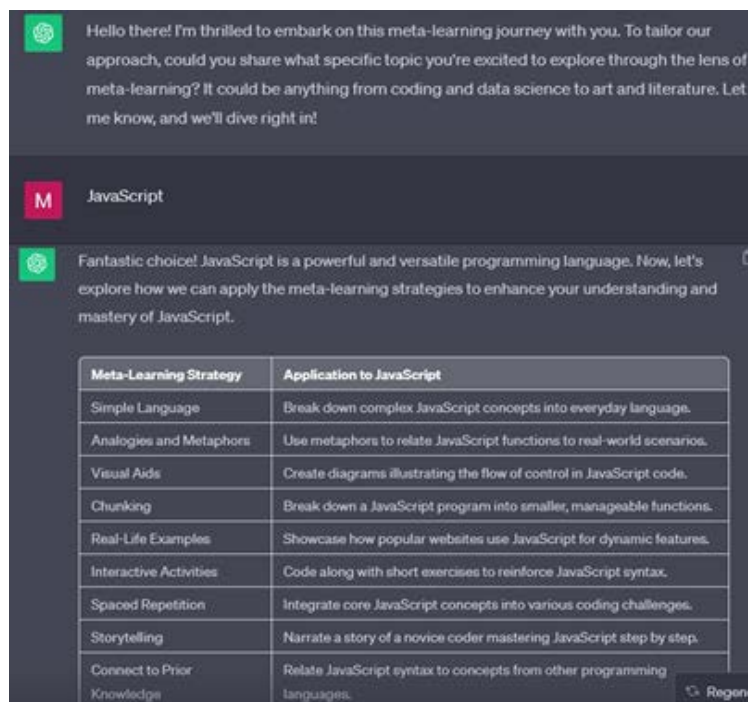


Figure 49: ChatGPT as a meta-learning expert proposing strategies for learning the JavaScript programming language.

²³ <https://flowgpt.com/p/learning-anything-the-power-of-meta-learning>

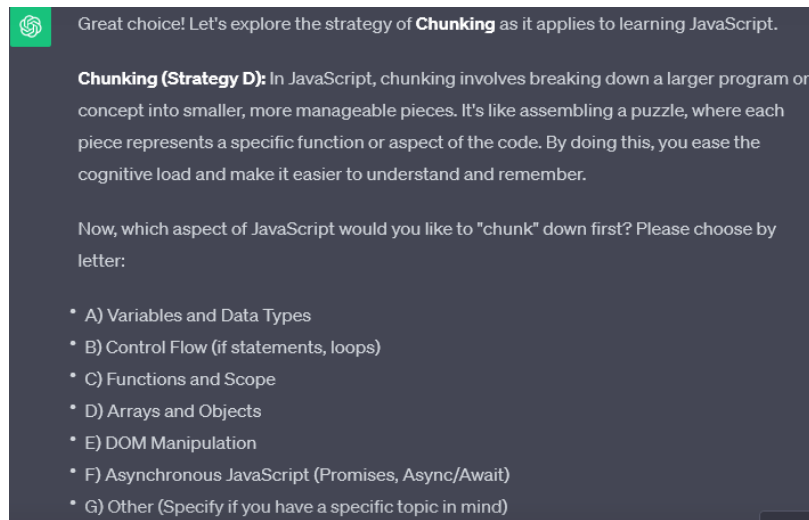


Figure 50: ChatGPT elaborates on the chunking meta-learning strategy based on JavaScript concepts.

3.5 Challenges

The integration of conversational AI in education has the potential to improve various areas in the educational field. However, studies have shown that the implications of its pedagogical adoption raise numerous concerns, including ethical, legal, and societal aspects [129]. Guidelines to mitigate these issues have also been proposed to establish a safer and more responsible implementation of these technologies.

Ethical aspects encompass user privacy and security issues, highlighting the importance of protecting users' personal and confidential information from being compromised and used for non-educational purposes [129]. It is becoming obvious that data privacy and security policies should comply with regulatory requirements (e.g., GDPR, FERPA, and others) and ethical standards accompanied by robust tools that ensure safe and ethical data usage techniques (e.g., anonymized data and secure infrastructure with modern encryption). The stakeholders should also be informed in order to prevent effectively and tackle potential risks and share how data collection and usage is done by spreading awareness to students and their guardians.

Another ethical concern includes the possibility of promoting unfairness and bias between responses. Conversational AI applications that rely on LLMs are trained on massive data formulated and curated by humans, which holds the risk of generating and spreading biased and harmful content among educational communities, hindering their pedagogical utilization [36]. In combination with the obscurity surrounding the

underlying generative mechanisms of LLMs, matters are complicated even further since users cannot gain insight into the model's produced content [131]. Creating diverse training and fine-tuning datasets, along with proper pre-processing and post-processing techniques, can contribute to the prevention of these problems. Continuous upgrades and human evaluation are also critical to ensure the ability to comply with ethical values and appropriate policy frameworks [129].

In addition, the cost of adoption and maintenance of LLM-based chatbots may further develop inequity problems. LLM technologies can be expensive for educational institutions to monitor, maintain, and update. That raises concerns regarding the equal dissemination of knowledge across education communities worldwide, especially in cases of commercial product use [36]. Primary suggestions include the use of open-source fine-tuned models on particular educational tasks, as well as initiatives for cooperation with research institutions or governmental and nonprofit organizations. Hence, schools and educational institutions will be reinforced with economic assistance and expertise or with the use of robust, cost-efficient resources through collaborative, scalable technologies, such as cloud-based [129]. Additionally, the high computational demands of these technologies significantly impact energy costs. This issue can be tackled by embracing energy-saving hardware and shared facilities that rely on renewable energy sources. Lastly, to ensure sustainability in training and optimization, it is crucial to use data in accordance with ethical practices that can be enforced by governmental regulations. Research promotion on adopting more efficient solutions for training and maintenance is also warranted.

Usage issues are also important since the easiness of acquiring responses and information can lead to an overreliance on these systems, both from students and teachers [36]. That may lead to implications, such as adopting plagiarism tactics that risk academic integrity (e.g., submitting papers or answering exams with AI-generated text) or resorting to using such tools for every solution to their problems, thus progressively relinquishing their efforts for critical and independent thinking as well as problem-solving [129]. When Conversational AI is integrated into educational settings as a supportive and complementary learning tool rather than a simple generative machine, it holds the potential to motivate exploration, investigation, and sharpening of teaching and learning skills. Experimentation with these technologies is a useful way for students to comprehend the limitations and disadvantages of their

use. Especially when combined with multi-resource teaching strategies (e.g., collaborative, project-based learning), they are not only able to assess schoolwork to ensure originality but also to engage students in activities where they validate the accuracy of these technologies and develop and refine their critical thinking skills.

Closely related to the above, the educational staff may face challenges regarding the proper incorporation of these systems into lessons since adopting emerging technologies requires a good understanding of their strengths and weaknesses [129]. Suggestions to effectively face these challenges include professional training sessions and the creation of relevant guidelines, as well as learning communities that will provide valuable insight to education professionals on how to leverage chatbots in teaching practice and upskill their competencies to exploit their advantages [132].

Chatbots may produce inaccurate or generic responses since, as stated before, the nature of the output is related to the nature of the available data. In addition, the response resemblance to human text may lead users to receive distorted information misrepresented as true. Consequently, it becomes crucial for students to be guided into being skeptical, verify their resources, and be mindful of the validity of the output they receive to prevent the diffusion of erroneous information [36]. Another obstacle arising from the human-like chatbot output is the challenge for teachers to differentiate the AI-generated answers from those produced by students. In this case, several strategies have been proposed, including the adoption of tools that can identify AI-generated text, such as GPTZero²⁴ or even ChatGPT, which is capable of recognizing its own responses, as well as watermarking the content created by language models [129]. Ultimately, studies agree that it becomes necessary to promote further research aimed at the development of efficient methods and mechanisms capable of detecting human or model-produced replies, the creation of new assessment procedures, the fostering of critical thinking during courses, as well as the provision of guidance to learners on the productive and supplemental utilization of chatbots [129] [132].

Regarding copyright matters concerning the generated text of conversational AI tools, such as ChatGPT, the US Copyright Office has stated that a machine or mechanical program running randomly or automatically is not protected by copyright unless there is human-author interference [129]. As such, human creations are

²⁴ <https://gptzero.me/>

exclusively protected by copyright [36]. On the other hand, professionals who undertake the role of training LLM applications using educational resources should be aware of the copyright terms and conditions, asking for authors' permission to include their content for the training set while informing users about the terms of use regarding the model output.

Conclusions

The present thesis is an investigation of conversational AI technologies, their application, and their impact on the education sector. It illuminates the milestones in the development of conversational agents or chatbots, uncovering their historical and technological trajectory that serves as a springboard for delving into the intricate world of AI in chatbot development. The technological advancements in DL and the leading role of NLP, especially language modeling, are underscored in shaping the modern conversational AI path. The game-changing LLMs are exemplified by ChatGPT, as a significantly impactful case of a chatbot powered by this technology. Lastly, the focus shifts towards the application of conversational AI in educational contexts. It investigates the catalytic role of AI chatbots as an instantiation of the integration of emerging Information and Communication Technologies (ICTs) in education through numerous educational use cases for the creation of interactive and personalized learning experiences. The challenges inherent in this integration process are also highlighted, including primary recommendations for tackling these issues.

The exploration of use cases demonstrates the transformative and reinventing potential of conversational AI, as stated by previous research. It offers plentiful and innovative opportunities in various educational aspects, which benefit and enhance the pedagogical environments with engaging, adaptable, and inclusive pedagogical approaches that the contemporary education era requires. Toward this direction, educators are encouraged to enhance their instructional skills and reorient their teaching practices to understand and exploit the strengths of these systems. In addition, policymakers and technologists should contribute to the refinement and navigation of the evolving landscape of conversational AI in education.

The challenges highlight the need for research promotion to holistically explore the various aspects of these technologies and ensure transparent and responsible AI deployment. This will promote fair and equal access to each individual, which is imperative before these systems are fully integrated into the education sector.

This work aims to contribute to and encourage the ever-growing research and discussion concerning the utilization of chatbots in education, especially of recent LLM-enabled systems that few studies have addressed. Future research directions of

this study can concentrate on the real-world educational application of LLM-based chatbots to evaluate their performance and difficulties. Additionally, the long-term learning outcomes and students' attitudes can be studied, encompassing factors such as engagement and motivation to use these tools.

References

- 1] [B. e. a. Khosrawi-Rad, "Conversational Agents in Education – A Systematic Literature Review," in *European Conference on Information Systems*, Timișoara, Romania, 2022.
- 2] [G.-J. Hwang and C.-Y. Chang, "A review of opportunities and challenges of chatbots in education," *Interactive Learning Environments*, vol. 31, no. 7, pp. 4099-4112, 2021.
- 3] [C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," *Computers and Education: Artificial Intelligence*, vol. 2, 2021.
- 4] [J. Kumar, "Educational chatbots for project-based learning: investigating learning outcomes for a team-based design course," *Int J Educ Technol High Educ.*, vol. 18, no. 1, 2021.
- 5] [S. Laato, B. Morschheuser, J. Hamari and B. J., "AI-Assisted Learning with ChatGPT and Large Language Models: Implications for Higher Education," in *IEEE International Conference on Advanced Learning Technologies (ICALT)*, Orem, UT, USA, 2023.
- 6] [E. Ruane, A. Birhane and A. Ventresque, *Conversational AI: Social and Ethical Considerations*, 2019.
- 7] [P. e. a. Kulkarni, "Conversational AI: An Overview of Methodologies, Applications & Future Scope," in *5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, Pune, India, 2019.
- 8] [M. McTear, *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots*, Morgan & Claypool Publishers, 2021.
- 9] [M. McTear, Z. Callejas and D. Griol, *The Conversational Interface: Talking to Smart Devices*, Springer Cham, 2016.
- 10] [M. Kuhail, N. Alturki and S. e. a. Alramlawi, "Interacting with educational chatbots: A systematic review," *Educ Inf Technol*, vol. 28, p. 973–1018, 2023.
- 11] [M. Allouch, A. Azaria and R. Azoulay, "Conversational Agents: Goals, Technologies, Vision and Challenges," *Sensors*, vol. 21, p. 8448, 2021.
- 12] [M. Wahde and M. Virgolin, *Conversational Agents: Theory and Applications*, 2022.
- 13] [E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Machine Learning with Applications*, vol. 2, 2020.
- 14] [G. Caldarini, S. Jaf and K. McGarry, "A Literature Survey of Recent Advances in Chatbots," *Information*, vol. 13, no. 1, p. 41, 2022.
- 15] [R. Agarwal and M. Wadhwa, "Review of State-of-the-Art Design Techniques for Chatbots," *SN Computer Science*, vol. 1, p. 246, 2020.
- 16] [D. Jurafsky and J. H. Martin, "Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd ed. draft)," Under development, 2023.
- 17] [R. Carpenter, "About the Jabberwacky AI," [Online]. Available: <http://www.jabberwacky.com/j2about>. [Accessed 20 May 2023].
- 18] [Wikipedia, "Cleverbot," Wikimedia Foundation, 20 August 2023. [Online]. Available: <https://en.wikipedia.org/wiki/Cleverbot>. [Accessed 20 September 2023].
- 19] [M. L. Mauldin, "CHATTERBOTS, TINYMUDS, and the Turing Test Entering the Loebner Prize Competition," in *AAAI'94: Proceedings of the Twelfth AAAI National*

- Conference on Artificial Intelligence*, Seattle Washington, 1994.
- [W. contributors, "Dr. Sbaitso," Wikipedia, The Free Encyclopedia, 24 April 2023.
20] [Online]. Available:
https://en.wikipedia.org/w/index.php?title=Dr._Sbaitso&oldid=1151578068.
[Accessed 16 May 2023].
- [Boost.ai, "What is a chatbot?," [Online]. Available: <https://www.boost.ai/chatbot>.
21] [Accessed 18 September 2023].
- [E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," in *IFIP*
22] *International Conference on Artificial Intelligence Applications and Innovations*, 2020.
- [N. Zierau, E. Elshan, C. Visini and A. Janson, "A Review of the Empirical Literature
23] on Conversational Agents and Future Research Directions," in *International Conference on Information Systems (ICIS)*, 2020.
- [Microsoft Asia, "Much more than a chatbot: China's Xiaoice mixes AI with
24] emotions and wins over millions of fans," 1 November 2018. [Online]. Available:
<https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-mixes-ai-with-emotions-and-wins-over-millions-of-fans/>. [Accessed 2023 April 10].
- [S. Hussain, A. O. Sianaki and N. Ababneh, "A Survey on Conversational
25] Agents/Chatbots Classification and Design Techniques," in *WAINA: Workshops of the International Conference on Advanced Information Networking and Applications*, Matsue, Japan, 2019.
- [C.-C. Lin, A. Y. Q. Huang and Y. S. J. H., "A Review of AI-Driven Conversational
26] Chatbots Implementation Methodologies and Challenges (1999–2022)," *Sustainability*, vol. 15, no. 15, p. 4012, 2023.
- [W. contributors, "ELIZA," 2023. [Online]. Available:
27] <https://en.wikipedia.org/wiki/ELIZA>. [Accessed 24 April 2023].
- [S. Meshram, N. Naik, M. VR, T. More and S. Kharche, "Conversational AI:
28] Chatbots," in *International Conference on Intelligent Technologies (CONIT)*, Hubli, India, 2021.
- [N. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating
29] Your First Ontology," Stanford Knowledge Systems Laboratory, 2001.
- [E. e. a. Khosla, "RNNs-RT: Flood based Prediction of Human and Animal deaths in
30] Bihar using Recurrent Neural Networks and Regression Techniques," *Procedia Computer Science*, vol. 132, pp. 486-497, 2018.
- [X. H. e. a. Le, "Application of Long Short-Term Memory (LSTM) Neural Network for
31] Flood Forecasting," *Water*, vol. 11, p. 1387, 2019.
- [L. Zhou, J. Gao, D. Li and H.-Y. Shum, "The Design and Implementation of Xiaoice,
32] an Empathetic Social Chatbot," *Computational Linguistics*, vol. 46, no. 1, pp. 53-93, 2020.
- [A. Shaikh, D. More, R. Puttoo, S. Shrivastav and S. Shinde, "A Survey Paper on
33] Chatbots," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 4, April 2019.
- [T. e. a. Shi, "Neural Abstractive Text Summarization with Sequence-to-Sequence
34] Models," *ACM/IMS Trans. Data Sci.*, vol. 2, no. 1, p. 37, 2020.
- [A. e. a. Vaswani, "Attention is All you Need," in *Neural Information Processing*
35] *Systems*, 2017.
- [C. e. a. Zhang, "One Small Step for Generative AI, One Giant Leap for AGI: A
36] Complete Survey on ChatGPT in AIGC Era," *ArXiv*, 2023.

- [37] H. V. P. Singh and Q. Mahmoud, "NLP-Based Approach for Predicting HMI State Sequences Towards Monitoring Operator Situational Awareness," *Sensors*, vol. 20, p. 3228, 2020.
- [38] F. Karami, "Decoding the Magic of Self-Attention: A Deep Dive into its Intuition and Mechanisms," 10 July 2023. [Online]. Available: <https://medium.com/@farzad.karami/decoding-the-magic-of-self-attention-a-deep-dive-into-its-intuition-and-mechanisms-394aa98f34c5>. [Accessed 16 August 2023].
- [39] J. Alammar, "The Illustrated Transformer," 27 June 2018. [Online]. Available: <http://jalammar.github.io/illustrated-transformer/>. [Accessed 10 August 2023].
- [40] W. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J. Nie and J. Wen, "A Survey of Large Language Models," 2023.
- [41] H. Wang, J. Li, H. Wu, E. Hovy and Y. Sun, "Pre-Trained Language Models and Their Applications," *Engineering*, 2022.
- [42] B. e. a. Min, "Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey," *ACM Computing Surveys*, 2023.
- [43] L. Hang, "Language Models: Past, Present, and Future," *Communications of the ACM*, vol. 65, no. 7, 2022.
- [44] D. Hiemstra, "Language Models," in *Encyclopedia of Database Systems*, L. LIU and M. ÖZSU, Eds., Springer, Boston, MA, 2009, pp. CCCXLV, 4355.
- [45] J. e. a. Bengio, "A Neural Probabilistic Language Model," *Journal of Machine Learning Research*, vol. 3, pp. 1137-1155, 2003.
- [46] D. & M. J. Jurafsky, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2008.
- [47] Turing, "A Guide on Word Embeddings in NLP," [Online]. Available: <https://www.turing.com/kb/guide-on-word-embeddings-in-nlp>. [Accessed 12 7 2023].
- [48] B. Luo, R. Y. K. Lau and C. & S. Y.-W. Li, "A critical review of state-of-the-art chatbot designs and applications," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(1), e1434, vol. 12, no. 1, 2022.
- [49] N. Bensalah, H. Ayad, A. Adib and A. Ibn El Farouk, "Combining Word and Character Embeddings for Arabic Chatbots," in *Advanced Intelligent Systems for Sustainable Development (AI2SD'2020)*, Tangier, Morocco, 2022.
- [50] Z. e. a. Kastrati, "WET: Word Embedding-Topic distribution vectors for MOOC video lectures dataset," *Data in Brief.*, vol. 28, p. 105090, 2020.
- [51] J. Li, T. Tang, W. Zhao and J.-R. Wen, "Pretrained Language Models for Text Generation: A Survey," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, Montreal, 2021.
- [52] M. e. a. Zhong, "DIALOGLM: Pre-trained Model for Long Dialogue Understanding and Summarization," 2022.
- [53] Y. e. a. Zhang, "DIALOGPT: Large-Scale Generative Pre-training for Conversational Response Generation," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Online, "Association for Computational Linguistics, 2020, pp. 270-278.
- [54] S. e. a. Roller, "Recipes for building an open-domain chatbot," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, Online, Association for Computational Linguistics, 2021, pp. 300-325.

- [55] R. e. a. Thoppilan, "LaMDA: Language Models for Dialog Applications," *arXiv preprint arXiv:2201.08239*, 2022.
- [56] S. e. a. Bae, "Building a Role Specified Open-Domain Dialogue System Leveraging Large-Scale Language Models," 2022.
- [57] M. Zaib, Q. Sheng and W. Zhang, "A Short Survey of Pre-trained Language Models for Conversational AI-A New Age in NLP," in *ACSW '20: Proceedings of the Australasian Computer Science Week Multiconference*, Melbourne, VIC, Australia, Association for Computing Machinery, 2020, pp. 1-4.
- [58] J. e. a. Wei, "Emergent abilities of large language models," *Transactions on Machine Learning Research*, 2022.
- [59] O. Honovich, T. Scialom, O. Levy and T. Schick, "Unnatural Instructions: Tuning Language Models with (Almost) No Human Labor," 2022.
- [60] J. Wei and D. Zhou, "Language Models Perform Reasoning via Chain of Thought," Google Research, Brain team, 2022.
- [61] O. AI, "OpenAI Charter," 9 April 2018. [Online]. Available: <https://openai.com/charter>. [Accessed 19 September 2023].
- [62] D. Mhlanga, "Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning," in *FinTech and Artificial Intelligence for Sustainable Development. Sustainable Development Goals Series*, Palgrave Macmillan, Cham, 2023.
- [63] A. Radford and K. Narasimhan, "Improving Language Understanding by Generative Pre-Training," 2018.
- [64] A. e. a. Radford, "Language Models are Unsupervised Multitask Learners," 2019.
- [65] S. e. a. Min, "Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?," in *EMNLP*, 2022.
- [66] J. Kossen, T. Rainforth and Y. Gal, "In-Context Learning in Large Language Models Learns Label Relationships but Is Not Conventional Learning," *ArXiv [cs.CL]*, 2023.
- [67] OpenAI, "GPT-4 Technical Report," 2023.
- [68] J. Rudolph, S. Tan and S. Tan, "War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education," *Journal of Applied Learning and Teaching*, vol. 6, no. 1, 2023.
- [69] OpenAI, "Introducing ChatGPT," OpenAI, 30 November 2022. [Online]. Available: <https://openai.com/blog/chatgpt>. [Accessed 24 July 2023].
- [70] A. Haleem, M. Javaid and R. P. Singh, "An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 2, no. 4, 2022.
- [71] A. B. Kocaballi, "Conversational AI-Powered Design: ChatGPT as Designer, User, and Product," 2023.
- [72] J. e. a. White, "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT," 2023.
- [73] OpenAI, "Custom instructions for ChatGPT," 20 July 2023. [Online]. Available: <https://openai.com/blog/custom-instructions-for-chatgpt>.
- [74] V. L. Tinio, *ICT in education*, Manila, Philippines: E-ASEAN Task Force, 2003.
- [75] A. F. a. A. O. (. o. t. U. Nations, "Information and Communication Technologies (ICT)," [Online]. Available: <https://aims.fao.org/information-and-communication->

- technologies-ict. [Accessed 18 September 2023].
- [P. Limna, S. Jakwatanatham, S. Siripipattanakul, P. Kaewpuang and P. 76] Sriboonruang, "A Review of Artificial Intelligence (AI) in Education during the Digital Era," *Advance Knowledge for Executives*, vol. 1, no. 1, pp. 1-9, 2022.
- [D. Amutha, "The Role and Impact of ICT in Improving the Quality of Education," 77] 2020.
- [J. S. Fu, "ICT in Education: A Critical Literature Review and Its Implications," 78] *International Journal of Education and Development using Information and Communication Technology (IJEDICT)*, vol. 9, no. 1, pp. 112-125, 2013.
- [J. Dron, "Educational technology: what it is and how it works," *AI & Soc*, vol. 37, 79] pp. 155-166, 2022.
- [M. e. a. Jhonattan, "The core components of education 4.0 in higher education: 80] Three case studies in engineering education," *Computers & Electrical Engineering*, vol. 93, no. 107278, 2021.
- [M. Ramírez-Montoya and L. I. González-Pérez, "Components of Education 4.0 in 81] 21st Century Skills Frameworks: Systematic Review," *Sustainability*, vol. 14, no. 3, p. 1493, 2022.
- [E. Mukul and G. Büyüközkan, "Digital transformation in education: A systematic 82] review of education 4.0," *Technological Forecasting and Social Change*, vol. 194, 2023.
- [E. Dimitriadou and A. Lianitis, "A critical evaluation, challenges, and future 83] perspectives of using artificial intelligence and emerging technologies in smart classrooms," *Smart Learn. Environ.*, vol. 10, no. 12, 2023.
- [D. Rotolo, D. Hicks and B. Martin, "What Is an Emerging Technology?," 2016. 84]
- [G. Veletsianos, "The Defining Characteristics of Emerging Technologies and 85] Emerging Practices in Digital Education," in *Emergence and Innovation in Digital Learning: Foundations and Applications*, Athabasca University Press, 2016, p. 227.
- [V. e. a. Mechelen, "Emerging Technologies in K–12 Education: A Future HCI 86] Research Agenda," *ACM Transactions on Computer-Human Interaction*, vol. 30, no. 3, pp. 1-40, 2023.
- [C. Cacciuttolo, Y. Vásquez, D. Cano and F. Valenzuela, "Research Thesis for 87] Undergraduate Engineering Programs in the Digitalization Era: Learning Strategies and Responsible Research Conduct Road to a University Education 4.0 Paradigm," *Sustainability*, vol. 15, no. 14, 11206, 2023.
- [S. Wollny, J. Schneider, D. Di Mitri, J. Weidlich, M. Rittberger and D. Hendrik, "Are 88] We There Yet? - A Systematic Literature Review on Chatbots in Education," *Front. Artif. Intell.*, vol. 4, 2021.
- [C. Kooli, "Chatbots in Education and Research: A Critical Examination of Ethical 89] Implications and Solutions," *Sustainability*, vol. 15, no. 7, p. 5614, 2023.
- [K. K. Mageira, D. Pittou, A. Papasalouros, K. Kotis, P. Zangogianni and A. 90] Daradoumis, "Educational AI Chatbots for Content and Language Integrated Learning," *Applies Sciences*, vol. 12, no. 7, p. 3239, 2022.
- [M. e. a. Aleedy, "Using AI Chatbots in Education: Recent Advances Challenges and 91] Use Case," in *Artificial Intelligence and Sustainable Computing, Algorithms for Intelligent Systems*, M. Pandit, M. Gaur, P. Rana and A. Tiwari, Eds., Springer, Singapore, 2022.
- [T. e. a. Wambsganß, "ArgueTutor: An Adaptive Dialog-Based Learning System for

- 92] Argumentation Skills," pp. 1-13, 2021.
- [A. Deveci Topal, C. Dilek Eren and A. Kolburan Geçer, "Chatbot application in a 5th
93] grade science course," *Educ Inf Technol*, vol. 26, p. 6241–6265, 2021.
- [Y. V. Lakshmi and I. Majid, "Chatbots in Education System," *University News: A
94] Weekly Journal of Higher Education*, vol. 60, no. 8, pp. 15-18, 2022.
- [M. Lin and H. Chang, "Enhancing Post-secondary Writers' Writing Skills with a
95] Chatbot: A Mixed-Method Classroom Study," *J. Educ. Technol. Soc.*, vol. 23, pp. 78-92,
2020.
- [J. F.-R. M. O.-A. S. e. a. Pereira, "Promoting Learners' Voice Productions Using
96] Chatbots as a Tool for Improving the Learning Process in a MOOC," *Tech Know Learn*,
vol. 24, p. 545–565, 2019.
- [J. Yin, T.-T. Goh, B. Yang and Y. Xiaobin, "Conversation Technology With Micro-
97] Learning: The Impact of Chatbot-Based Learning on Students' Learning Motivation and
Performance," *Journal of Educational Computing Research*, vol. 59, no. 1, pp. 154-177,
2021.
- [D. Jackson and A. Latham, "Talk to The Ghost: The Storybox methodology for
98] faster development of storytelling chatbots," *Expert Systems with Applications*, vol.
190, 2022.
- [M. M. Sanchez, A. C. Melo, L. S. Blanco and A. M. F. García, "Chatbot, as
99] Educational and Inclusive Tool for People with Intellectual Disabilities," *Sustainability*,
vol. 14, no. 3, p. 1520, 2022.
- [Y. Xu, V. Vigil, A. S. Bustamante and M. Warschauer, "'Elinor's Talking to Me!':
100] Integrating Conversational AI into Children's Narrative Science Programming," 2022.
- [C. Liu, M. Liao, C. Chang and H. Lin, "An analysis of children' interaction with an AI
101] chatbot and its impact on their interest in reading," *Computers & Education*, vol. 189,
2022.
- [D. Lee and S. Yeo, "Developing an AI-based chatbot for practicing responsive
102] teaching in mathematics," *Computers & Education*, vol. 191, 2022.
- [K. Na-Young, "A Study on the Use of Artificial Intelligence Chatbots for Improving
103] English Grammar Skills," *Journal of Digital Convergence*, vol. 17, pp. 37-46, 2019.
- [G. Terzopoulos and M. Satratzemi, "Voice Assistants and Artificial Intelligence in
104] Education," in *Proceedings of 9th Balkan Conference on Informatics (BCI'19)*, Sofia,
Bulgaria, 2019.
- [M. Bao, "Can Home Use of Speech-Enabled Artificial Intelligence Mitigate Foreign
105] Language Anxiety – Investigation of a Concept," *Arab World English Journal (AWEJ)
Special Issue on CALL*, vol. 5, pp. 28-40, 2019.
- [I. Hapsari and T. Wu, "AI Chatbots Learning Model in English Speaking Skill:
106] Alleviating Speaking Anxiety, Boosting Enjoyment, and Fostering Critical Thinking," in
Innovative Technologies and Learning. ICITL 2022. Lecture Notes in Computer Science,
2022.
- [Y. Ye, J. Deng and X. Liu, "The Effect of Oral Practice via Chatbot on Students' Oral
107] English Accuracy," in *International Symposium on Emerging Technologies for
Education. SETE 2021. Lecture Notes in Computer Science()*, 2021.
- [S. e. a. Ruan, "BookBuddy: Turning Digital Materials Into Interactive Foreign
108] Language Lessons Through a Voice Chatbot," 2019.
- [S. Mendoza, M. Hernández-León, L. Sánchez-Adame, J. Rodríguez, D. Decouchant
109] and A. Meneses-Viveros, "Supporting Student-Teacher Interaction Through a Chatbot,"

- in *Learning and Collaboration Technologies. Human and Technology Ecosystems. HCI 2020. Lecture Notes in Computer Science()*, 2020.
- [S. Sinha, S. Basak, Y. Dey and A. Mondal, "An Educational Chatbot for Answering
110] Queries," in *Emerging Technology in Modelling and Graphics. Advances in Intelligent Systems and Computing*, 2020.
- [A. K. Goel and L. Polepeddi, "Jill Watson: A Virtual Teaching Assistant for Online
111] Education," 2016.
- [A. Goel, "AI-Powered Learning: Making Education Accessible, Affordable, and
112] Achievable," *ArXiv*, vol. abs/2006.01908, 2020.
- [S. Gupta and Y. Chen, "Supporting Inclusive Learning Using Chatbots? A Chatbot-
113] Led Interview Study," *Journal of Information Systems Education (JISE)*, vol. 33, no. 1, pp. 98-108, 2022.
- [S. e. a. Meshram, "College Enquiry Chatbot using Rasa Framework," in *Asian
114] Conference on Innovation in Technology (ASIANCON)*, Pune, India, 2021.
- [T. Kita, C. Nagaoka, N. Hiraoka and M. Dougiamas, "Implementation of Voice User
115] Interfaces to Enhance Users' Activities on Moodle," in *4th International Conference on Information Technology (InCIT)*, Bangkok, Thailand, 2019.
- [Y.-C. Lee and W.-T. Fu, "Supporting Peer Assessment in Education with
116] Conversational Agents," in *In 24th International Conference on Intelligent User Interfaces (IUI '19 Companion)*, Marina del Ray, CA, USA, 2019.
- [SURF, "Promises of AI in Education- Discussing the impact of AI systems in
117] educational practices," 2022.
- [A. S. e. al., "Mental health and academic performance: a study on selection and
118] causation effects from childhood to early adulthood," in *Soc Psychiatry Psychiatr Epidemiol.*, 2021.
- [M. C. Pascoe, S. E. Hetrick and A. G. Parker, "The impact of stress on students in
119] secondary school and higher education," *International Journal of Adolescence and Youth*, vol. 25, no. 1, pp. 104-112, 2019.
- [B. Ouatu and D. Gifu, "Chatbot, the Future of Learning?," in *Ludic, Co-design and
120] Tools Supporting Smart Learning Ecosystems and Smart Education. Smart Innovation, Systems and Technologies*, 2021.
- [G. D'Silva, M. Jani, V. Jadhav, A. Bhoir and P. Amin, "Career Counselling Chatbot
121] Using Cognitive Science and Artificial Intelligence," in *Advanced Computing Technologies and Applications. Algorithms for Intelligent Systems*, 2020.
- [N. e. a. Suresh, "Career Counseling Chatbot on Facebook Messenger Using AI," in
122] *Proceedings of the International Conference on Data Science, Machine Learning and Artificial Intelligence*, Windhoek, Namibia, 2022.
- [S. e. a. Demetriadis, "Conversational agents in MOOCs: reflections on first
123] outcomes of the colMOOC project," in *Intelligent Data-Centric Systems, Intelligent Systems and Learning Data Analytics in Online Education*, S. Caballé, S. N. Demetriadis, E. Gómez-Sánchez, P. M. Papadopoulos and A. Weinberger, Eds., Academic Press, 2021.
- [S. e. a. Tegos, "Agent-Supported Peer Collaboration in MOOCs," *Frontiers in
124] Artificial Intelligence*, vol. 4, 2021.
- [A. T. e. a. Neumann, "Chatbots as a Tool to Scale Mentoring Processes: Individually
125] Supporting Self-Study in Higher Education," *Frontiers in Artificial Intelligence*, vol. 4, 2021.

- [126] Woebot Health, "Relational Agent for Mental Health | Woebot Health," [Online]. Available: <https://woebothealth.com/>. [Accessed 6 October 2023].
- [127] C. Chun Ho, H. L. L. W. K. Lee and K. F. A. Lui, "Developing a Chatbot for College Student Programme Advisement," in *International Symposium on Educational Technology (ISET)*, Osaka, Japan, 2018.
- [128] J. Maldonado-Mahauad, M. Pérez-Sanagustín, J. Carvallo-Vega, E. Narvaez and M. Calle, "Miranda: A Chatbot for Supporting Self-regulated Learning," in *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption. EC-TEL 2022. Lecture Notes in Computer Science*, 2022.
- [129] E. e. a. Kasneci, "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learning and Individual Differences*, vol. 103, 2023.
- [130] M. Sallam, "The Utility of ChatGPT as an Example of Large Language Models in Healthcare Education, Research and Practice: Systematic Review on the Future Perspectives and Potential Limitations," *Healthcare*, vol. 11, no. 6, p. 887, March 2023.
- [131] N. Ahmad, S. Murugesan and N. Kshetri, "Generative Artificial Intelligence and the Education Sector," *Computer*, vol. 56, no. 6, pp. 72-76, June 2023.
- [132] A. e. a. Tlili, "What if the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education," *Smart Learning Environments*, vol. 10, no. 1, pp. 1-24, 2023.
- [134] N. E. Evangelopoulos, "Latent semantic analysis," *WIREs Cognitive Science*, vol. 4, no. 6, pp. 683-692, 2013.
- [135] A. P. Association, "APA Dictionary of Psychology," [Online]. Available: <https://dictionary.apa.org/finite-state-grammar>. [Accessed 14 August 2023].