



# BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

## FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY  
A KOMUNIKAČNÍCH TECHNOLOGIÍ

## DEPARTMENT OF FOREIGN LANGUAGES

ÚSTAV JAZYKŮ

## QUALITATIVE ANALYSIS OF PROFESSIONAL TERMINOLOGY IN SPECIALIZED FIELDS

KVALITATIVNÍ ANALÝZA ODBORNÉ TERMINOLOGIE SPECIALIZOVANÝCH OBORŮ

### BACHELOR'S THESIS

BAKALÁŘSKÁ PRÁCE

#### AUTHOR

AUTOR PRÁCE

Bc. Adam Homolka

#### SUPERVISOR

VEDOUCÍ PRÁCE

PhDr. Bc. Dagmar Šťastná

BRNO 2023

# Bachelor's Thesis

Bachelor's study field **English in Electrical Engineering and Informatics**

Department of Foreign Languages

**Student:** Bc. Adam Homolka

**ID:** 228772

**Year of  
study:** 3

**Academic year:** 2022/23

## TITLE OF THESIS:

### **Qualitative analysis of professional terminology in specialized fields**

#### INSTRUCTION:

The aim of the bachelor thesis is to characterize the style of science and technology (features, functions, form, syntax, lexicon) and summarize the theoretical findings concerning translation of professional terminology and translation procedures at the lexical and grammatical levels. The practical part of the bachelor thesis will focus on the application of theoretical knowledge within the qualitative analysis of selected professional texts.

#### RECOMMENDED LITERATURE:

KNITTLOVÁ, D., GRYGOVÁ, B. & ZEHNALOVÁ, J. Překlad a překládání. Olomouc: Univerzita Palackého v Olomouci, 2010. ISBN 978-80-244-2428-6.

ŽVÁČEK, D. Kapitoly z teorie překlada I. (odborný překlad). Olomouc: Univerzita Palackého v Olomouci, 1995. ISBN 80-7067-489-x.

TAMMERMAN, R. Towards new ways of terminology description. Amsterdam/Philadelphia: John Benjamins Co., 2000.

**Date of project  
specification:** 9.2.2023

**Deadline for  
submission:** 30.5.2023

**Supervisor:** PhDr. Bc. Dagmar Šťastná

**doc. PhDr. Milena Krhutová, Ph.D.**  
Subject Council chairman

#### WARNING:

The author of the Bachelor's Thesis claims that by creating this thesis he/she did not infringe the rights of third persons and the personal and/or property rights of third persons were not subjected to derogatory treatment. The author is fully aware of the legal consequences of an infringement of provisions as per Section 11 and following of Act No 121/2000 Coll. on copyright and rights related to copyright and on amendments to some other laws (the Copyright Act) in the wording of subsequent directives including the possible criminal consequences as resulting from provisions of Part 2, Chapter VI, Article 4 of Criminal Code 40/2009 Coll.

## **Abstrakt**

Cílem bakalářské práce je charakterizovat odborný styl (rysy, funkce, forma, syntax, lexikum) a následně shrnout teoretické poznatky z oblasti překladu odborného stylu týkající se terminologie a překladatelských postupů na úrovni lexikální a gramatické. Praktická část bakalářské práce se zaměří na aplikaci teoretických poznatků v rámci kvalitativní analýzy vybraných odborných textů.

**Klíčová slova:** Odborný styl, rysy, funkce, forma, syntax, lexikum, terminologie, lexikologie, překlad, kvalitativní analýza, analýza lexikální, analýza syntaktická, analýza textová, analýza pragmatická

## **Abstract**

The aim of the Bachelor's thesis is to characterize professional style (features, functions, form, syntax, lexicon) and then to summarize the theoretical knowledge in the field of translation of professional style concerning terminology and translation procedures at the lexical and grammatical level. The practical part of the bachelor's thesis will focus on application of theoretical knowledge in the quantitative analysis of selected professional texts.

**Key words:** Professional style, features, functions, form, syntax, lexicon, terminology, lexicology, translation, qualitative analysis, lexical analysis, syntax analysis, text analysis, pragmatic analysis

## Author's Declaration

**Author:** *Adam Homolka*

**Author's ID:** *228772*

**Paper type:** *Bachelor's thesis*

**Academic year:** *2022/2023*

**Topic:** *Qualitative analysis of professional terminology in specialized fields*

I declare that I have written this paper independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the paper and listed in the comprehensive bibliography at the end of the paper.

As the author, I furthermore declare that, with respect to the creation of this paper, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation S 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll. of the Czech Republic, Section 2, Head VI, Part 4.

Brno: 30.5.2023

-----  
Author's signature

## Acknowledgement

I would like to thank my thesis supervisor PhDr. Bc. Dagmar Šťastná for her support and valuable comments.

## Table of contents

Table of contents .....	5
List of abbreviations .....	8
List of tables .....	9
Introduction .....	10
1 Style of science and technology .....	11
1.1 Features .....	12
1.2 Functions .....	13
1.3 Forms .....	14
1.4 Lexicon .....	15
1.5 Syntax .....	15
1.6 Summary .....	16
2 Translation of professional texts .....	17
2.1 Translation procedures described by Vinay and Darbelnet .....	18
2.2 Translation procedures by Joseph L. Malone .....	19
2.3 Terminology .....	21
2.4 Lexicology .....	22
2.5 Difference between terminology and lexicology .....	23
2.6 Terms and their translation (general vs. specialized dictionaries) .....	23
2.6.1 General-purpose dictionary .....	23
2.6.2 Specialized dictionary .....	24
3 Research objectives and methodology .....	25
3.1 Objectives and research questions .....	25
3.2 Research materials .....	26

3.3	Aspects of the analysis .....	27
3.4	Procedure.....	28
3.5	Summary .....	29
4	Analysis.....	30
4.1	Text 1.....	30
4.1.1	Lexical level.....	30
4.1.2	Syntax level.....	32
4.1.3	Text level .....	33
4.1.4	Pragmatic level.....	34
4.1.5	Conclusion .....	35
4.2	Text 2.....	36
4.2.1	Lexical level.....	36
4.2.2	Syntax level.....	38
4.2.3	Text level .....	40
4.2.4	Pragmatic level.....	41
4.2.5	Conclusion .....	42
4.3	Text 3.....	43
4.3.1	Lexical level.....	43
4.3.2	Syntax level.....	44
4.3.3	Text level .....	47
4.3.4	Pragmatic level.....	48
4.3.5	Conclusion .....	49
4.4	Text 4.....	49
4.4.1	Lexical level.....	50

4.4.2	Syntax level .....	51
4.4.3	Text level .....	53
4.4.4	Pragmatic level .....	54
4.4.5	Conclusion.....	55
4.5	Comparison.....	56
4.5.1	Lexical level .....	56
4.5.2	Syntax level .....	58
4.5.3	Text level.....	59
4.5.4	Pragmatic level .....	60
	Conclusion.....	61
	List of references .....	64
	Extended abstract .....	67
	Appendices .....	70
	Article by The Guardian.....	70
	Article by Business Insider .....	73
	Article by BBC Science Focus.....	75
	Article generated by AI – Chat GPT.....	80



## List of abbreviations

AI – Artificial Intelligence

SL – Source Language

TL – Target Language

## List of tables

Table 1 Text 1 - The most frequent terms .....	30
Table 2 Text 1 - The most frequent terms in word classes.....	31
Table 3 Text 2 - The most frequent terms .....	36
Table 4 Text 2 - The most frequent terms in word classes.....	37
Table 5 Text 3 - The most frequent words .....	43
Table 6 Text 3 - The most frequent terms in word classes.....	44
Table 7 Text 4 - The most frequent terms .....	50
Table 8 Text 4 - The most frequent terms in word classes.....	50
Table 9 The most used words in all four texts .....	56
Table 10 The most used word classes in all four texts .....	56
Table 11 The most used nouns in all four texts.....	57
Table 12 Examples of professional adjectives in all four texts.....	58
Table 13 Examples of professional verbs in all four texts .....	58

## Introduction

The Bachelor thesis focuses on the qualitative analysis of professional texts. The theoretical part of the work will consist of an elementary description of the professional style and especially the style of science and technology. Firstly, I will describe its features, functions, forms, what lexical elements it uses, and specifications of its syntax. Secondly, it will describe translation of the professional text and procedures used to accomplish a successful translation. And lastly, the work will describe terminology principles.

The aim of the theoretical part is to introduce readers to the topic of the professional style of language. After reading this part, the practical part should be comprehensive for the reader. It will also be a knowledge base for me in my further practical part.

The practical part will be an analysis of four different texts concerned with the same topic. The analytical part will consist of four levels of analysis *lexical level*, *syntax level*, *text level*, and *pragmatic level*. The lexical level will analyze the vocabulary and language of the text, such as if the vocabulary matches the language of science and technology, if there are any internationalisms, or if the text is lexically dense. It will also analyze the most occurring word classes and words in each text. The syntax level will be an analysis of the structure within sentences, syntax, and overall coherence. The examples which will be observed in the syntax level are the complexity of sentences, types of sentences, or passive voice structures. The text level will deal with the visual structure of texts, such as whether the text is divided into paragraphs, visual division by headings, or what fonts were used. Lastly, in the pragmatic level, the analysis will contain examples of modality, situational context, author's intentions, and reader's expectations.

Some of the provided examples will be translated from English to Czech, and the translation will be analyzed.

The last part of the practical part will be a comparison in which all four analyses of all four texts will be compared. The aim is to find similarities and differences or analyze repeating patterns in the texts.

# 1 Style of science and technology

Style of science and technology is a type of professional style alongside with other fields of study or occupation – such as business, medicine, law, etc. According to Knittlová (1995) a professional style is an elementary style of non-fiction literature. Pastyřík (2010) says that the professional style has the intention to share information, analyses, and theoretical theses in the field of science, technology, and practice systematically and explicitly.

The language and style of professional texts were explicitly formed within the 19th century due to the industrial revolution. According to Čechová (2008), there was a substantial development in science and technology. Therefore, the scientific community needed to communicate information and ideas effectively. The process of language development has continued since then, given that technology has been still developing and introducing new scientific knowledge and fields of study. Krhutova (2007) points out the fact that the technical style has undergone rapid changes. The fact is based on the fact that technology, especially IT, has accelerated its development since the end of the 20th century, and the language used had to be adapted adequately.

Professional language is a broad topic, and it includes diverse types of professional texts, such as books, articles, encyclopedias, vocabularies, guides, textbooks, documentation, brochures, instructions, etc. Each of these professional texts is slightly different, which is determined stylistically. (Žvábek 1995)

Poštolková (1983) mentions that the professional style is one of the most frequently used styles – at least three-quarters of all written and printed production is written in the professional style. It is elementary for schools or university lectures, official documents, seminars, or non-fictional literature. The reader on which the text is focusing is predetermined. Both original and translated versions have the same information key, making their communication more accessible. Based on that, the text has a form with defined boundaries. (Man 2010)

Čechová, M., Krčmová, M and Minářová, E (2008) divide professional style into these substyles:

The first one is **scientific ( or theoretical)**, which contains new information or thoughts for the reader with the knowledge of a topic. The texts are well elaborated both syntax-wise and lexicon-wise. It fulfils international traditions and regulations of the field and contains citations, paraphrases, notes, bibliography, or an abstract. The second substyle is **practically professional**. It may not have to contain new information, however, the most crucial part is the practical use of the knowledge. The lexicon is not relatively complex, is less terminological, and sometimes informal. Thirdly there is **non-fictional** substyle. Such texts intend to share information and knowledge with the reader. Structurally, the syntax and lexicon are adjusted to the reader's need to be easily absorbed. The fourth one is **educative** in which the reader has only limited knowledge of the topic. The syntax and lexicon are customized to be easily perceivable. The author usually adds further additional information to make the text comprehensible and understandable. Lastly there is **essayistic** substyle which is subjective and represents a mixture of professional and fictional literature elements.

Plášek (2014) distinguishes specific procedures to elaborate a professional text. He mentions five techniques, which are as follows: informational, storytelling, descriptive, explanatory and reflective.

## 1.1 Features

The professional text must use symptomatic means of language and organize them in the context of the specific topic or field. The organization of the text is based on the logic and understanding of the meaning. Therefore, style is typical for its accuracy, unambiguity, and tightness of linguistic expression. (Hrdlička 2010)

According to Žvábek (1995), the most essential features of the professional style are the accuracy and consequentiality of the author's idea. He also specified the division of seven general features of the professional style which are: tightness of the syntax, terms saturation, abstract and non-agency expression, diversion from expressive means of

communication, minimizing the subjectivity of expressions, maximizing the use of terminology, and the necessity of precision and unambiguity of the expression.

The value of professional literature depends preeminently on the accuracy of the authors' description of the real world. The should be objective in the professional style. The authors' insights are objective and intended to be objectively shared with the reader (Hrdlička 2010). Čechová (2008) brings up **intertextuality** to refer appropriately to other literature through references, links, or notes. It is also crucial to preserve the text and its meaning to be understandable in the future. She mentions that the text must be composed logically with precise syntax, understanding international and domestic terminology and culture even in the lower levels of the language. Hence, according to Man (2010), the text must involve four elementary components – terminological, semi-special, scientific, and natural (spoken language).

Given these elements, the scientific one is considered the most interesting for a translator. The scientific component has given a semantic plan, and proportion of lexical features to expressed terms and the language is comprehensive and accurate based on codified automatizations. The intention is to provide a maximum parallel between a language form and expression semantics. In order to achieve it, the scientific language creates odd linguistic means which are unnecessary in other texts. (Čechová, 2003)

## 1.2 Functions

The text in the style of science and technology must be particularly specific because its most important function is to share information among concerned readers. The reader needs to have comprehension of the particular topic. If not, the text might be misunderstood or difficult to apprehend. As mentioned by Knittlová (2000), the main function is to convey information precisely and clearly to avoid any misunderstanding or ambiguity. The content of the text must be entirely explained and expressed. The text is usually in the form of a monologue with logical explanations and descriptions of specific terms related to the discussed topic. Those terms then linearly appear more often as the text becomes more specialized. In order to express information in the most comprehensive way for the reader, the text has to be stylistically clear due to its content complexity.

According to Čechová (2003), the essential function can be defined as a professional communication function. Consequently, it conveys a different function than the common language. The aim is to create an unambiguous and comprehensive idea about discussed matter in such a way that the readers may absorb provided information among their knowledge and subsequently create their own perspective to develop and further apply it. (Čechová, 2003)

### 1.3 Forms

As mentioned previously, the most frequently used written form is a monologue. The author uses a dialogue rarely in the scientific style. The professional style can take an oral form (such as in the lecture, report, or co-report), which is also mostly in the form of a monologue. Nevertheless, the oral form usually has a written template as the oral dialogue is rarely used (panel discussions). (Čechová, 2008)

It is also possible to divide the text horizontally and vertically. The horizontal division, which is also applied in this thesis, usually includes the items that are specified below:

The first example might be the introduction in which the reader is introduced to the thesis' structure and its content. It should be easy to understand even for a person without any professional knowledge. Another example is the aim – the aim should involve the reasoning why the topic is essential and what the author expects from the thesis' outcome. The theoretical part, which is a knowledge base and should involve all information used to elaborate the practical part, is generally divided horizontally into other chapters and subparts. The practical part should involve the process of research elaboration and is the most crucial part of the thesis. Similarly, as in the theoretical part, it is divided horizontally into other chapters and subparts. The discussion should highlight the value of the research and its comparison to the research thematically the same or similar. Last example is the conclusion which should summarize and evaluate the thesis. It might include recommendations describing how the research can be developed further or expanded, and ideas for designing the topic itself.

The vertical division consists of authors' notes, references to the used literature, citations, or paraphrases. The text itself is also structured by the font size and type into main and secondary thoughts.

Another type of division is by the style of writing and situational division. Forms divided by the style of writing are thesis, study, article, reflection, and essay. On the other hand, forms divided situationally are report, co-report, lecture, and discussion or debate. (Čechová, M. et al. 2008)

#### 1.4 Lexicon

According to Knittlová (1995), terms are the most crucial element of the professional style. The most frequent word class are **nouns/substantives**, usually professional unambiguous expressions. Both lexicon and syntax are stereotypical, which is adequate for the function of the style. To understand meaning comprehensively, the reader must acquire professional education. In order to justify the purpose of the text, the text predominantly consists of logical argumentation and facts. (Knittlová, 1995)

Čechová, M., et al. (2008) says that expressions used in professional texts are almost always formal and non-emotional/objective. If there is an oral speech in a professional style, it might non-intentionally contain informal speech due to inadequate preparation. In written form, informal elements might appear in non-fiction literature. Given the general meaning of the professional style, there is a considerable number of international terms and vocabulary. Moreover, part of the author's whole work is an abstract in multiple languages. (Čechová, M. et al, 2008)

#### 1.5 Syntax

As previously mentioned, the syntax is quite stereotypical in the professional style to correspond to the required form. Syntax usually consists of a variety of standard structures. The action originator is usually expressed vaguely, in general terms, or not at all. In addition, the author uses syntactic means such as impersonal sentences, sentences with general or indefinite subjects, and passive voice. (Hrdlička, 2010)



Professional texts have complicated structures consisting of long and complex sentences. That is a consequence of the complexity of the hierarchy of the author's thoughts which are presented in those texts. Given that, the text must have elements to help the reader to orientate in the complex text. To achieve better orientation, the author uses punctuation, brackets, or different types of fonts. In general, professional texts strive to make the number of statements as economical as possible, which means that the author is trying to share as much valuable information as possible in the lowest number of expressions possible. To achieve that goal, the author may exploit various infinitive constructions, agency nouns, adjectives, and transitive verbs. Simultaneously, the author uses established signs and symbols to achieve a more significant message condensation. (Čechová, M. et al., 2008)

## 1.6 Summary

In the first part of the chapter, the professional style and the style of science of technology was described and consequently, the professional style was divided into five substyles and mentioned five techniques to elaborate a professional text were mentioned. Subchapter 1.1 describes the features of the professional style and mentions the term intertextuality, and then the following subchapter deals with the functions of the professional style. Subchapter 1.3 describes the most common forms of such style and divides the text horizontally and vertically. Finally, subchapters 1.4 and 1.5 illustrate the lexicon and syntax of such style, so we can see what vocabulary is used and how the sentences are structured.

## 2 Translation of professional texts

In his book *Úvod do teorie překladau*, Levý (1958) specifies three stages of the translation process.

The first one is a process of **decoding** in which the translator attempts to comprehend the meaning and the intentions of the SL author. The translator then continues with a **coding** stage of the process in which translator analyses their findings in the SL and translates them to the TL without overshadowing the interpretation of the original author. The final stage is called a **reproduction**. There the translator uses a lexical equivalent without losing the semantic meaning.

Vinay and Darbelnet (1995) further specify the translation process itself. Translators establish relationships between two different linguistic systems where one of them (in the source language) has already been given and the second one (the target language) being potential and adaptable. Therefore, authors need to think of a final impression they want to create to the reader. To reach that, they can take initial steps. Authors should identify the units of translation, analyze descriptive, affective, and intellectual content of the SL text, reconstitute the situation which gave rise to the message, and lastly, weigh up and evaluate stylistic effects.

Knittlová (1995) then points out that translations used to be analyzed and reviewed as products in the past. She notices that nowadays, translators tend to focus on the process producing the final translation as a product. Monographies are often written in an unanimous form focusing on a macro-approach (cultural and historical background, structure of audience, or the type and function of the text). After the so-called strategic decision, translators implement a micro-approach focusing on particular elements to create a definite final text.

Translation is a process of transferring the written work in the source language to the written work in the target language whilst maintaining the meaning of content, aesthetics, style, and main idea of the original work. The translation is not just a statical replacement

of characters from one language to another, it is a complicated process requiring much skill in order to be successful. (Man, 2010)

As Hrdlička (2010) then adds, the main goal of the translation of the professional style is to capture the principal layer of meaning. Nevertheless, the accuracy of the translated text is not always achievable due to fact that even professional statements are not precisely formulated in all cases. Thus, some words are often expressed in non-terminological expressions.

The following subchapters deal with two main translating procedures where the first one was described by Vinay and Darbelnet whilst the second one was introduced by Joseph L. Malone.

## 2.1 Translation procedures described by Vinay and Darbelnet

The translation procedure described by Vinay and Darbelnet is the first classification written in 1958, and the techniques, which were introduced by the authors, are structured into seven classes:

First class is **borrowing**. Borrowing a process of taking the word from the source/original language (SL) and maintaining it in the target language (TL). The translated word is converted in a predictable manner. The process tends to be employed in two situations:

- I. In discussions about new technical processes without having any suitable term in the target language.
- II. To add a stylistic effect from the source language where the foreign term is added to flavour the target text.

An example of borrowing might be Czech word “robot”. No other language had an equivalent to this word; therefore, the word became widely used all around the world. It is also common in gastronomy where foods are not commonly translated (sushi, hamburger, wasabi, ...). The second class is a **calque**. The calque is a process of word-for-word translation of an expression from the source language into the target language. Calques either follow the syntax of the TL or maintain the syntax of the SL. For example, word

“skyscraper” is a word-for-word translation to Czech word “mrakodrap”. Czech language did not have a term for a very high building, therefore, it translated the English version word-for-word. The third class represents **literal translation**. It is a process of translating word for word in the way in which the word does not lose the meaning. It expands the scope of calque into more acceptable way. **Transposition** as the fourth class changes word class without changing the source meaning – such as the changes from nouns to verbs, adjectives to nouns, and vice versa. According to Vinay and Darbelnet, transposition is obligatory or optional, and it refers to the source text as the base expression and the target text as transposed expression. It is used to better fit the conventions of the of the TL. **Modulation** belongs to the fifth class and it is a process of modulating/rendering the TL from a different viewpoint than it is shown in the SL. It is necessary to avoid producing an awkward-sounding translation despite being grammatically, syntactically, and lexically correct. Moreover, the process of modulation finds other means or ways to express the idea to the target language without losing meaning from the original language. For example, a double negative is possible and quite common in the Czech language. However, in English, the negative is shown only once in the sentence. The sixth class involves an **equivalence** which is a process of using different stylistic means of the TL to describe the term in the SL. For example, a literal translation of an idiom in the SL would not make sense in the TL, so we need to use an equivalent idiom or phrase to describe the same idea. For example, an English idiom “to kick the bucket” means to die, however, literal translation would not make any sense in Czech, therefore, we use the term “natáhnout bačkory” which would mean “to pull up your slippers”. The last class, according to Vinay and Darbelnet, is an **adaptation**. The adaptation a process of adapting the idea in the SL to the TL whilst ensuring its relevance and meaningfulness as it is in the original. It would involve making significant changes to the SL text to make it more comprehending for a reader of the TL text. Examples might be cultural differences that need to be further specified in order to be fully understood.

## 2.2 Translation procedures by Joseph L. Malone

Thirty years later, in 1988, Joseph L. Malone supplemented the classification of Vinay and Darbelnet in his book *The science of linguistics in the art of translation* and divided 9

classes into pairs by their generic trajections – Matching (Equation and Substitution), Zigzagging (Divergence and Convergence), Recrescence (Amplification and Reduction), Repacking (Diffusion and Condensation), and Reordering which is unpaired. Those nine classes are further defined as it is illustrated in the following paragraphs:

The first class is the **equation**, which is a process of finding a perfect lexical equivalent in the TL for the term in the SL. The goal is to preserve the precise meaning of the SL. The next class is the **substitution** which is a process of replacing a term with a substitute element with the maximum semantic similarity if there is no lexical equivalent for the term. The goal is to find more naturally sounding expression in the TL. Another type is the **divergence**, which is a process of using more equivalents for a single term if there is more than one lexical equivalent. For example, in the Czech language, “ty/vy” is used to specify a relationship or closeness among people or plurality. However, there is only “you” in English for all previously mentioned options. The opposite of the divergence is the **convergence**. In the convergence is more than one equivalent in the SL but only one lexical equivalent in the TL. The translator may need to use different sources to create a coherent translation. Following the convergence is the **amplification**, which is a process of using further description in the TL in addition to the lexical equivalent in order to improve the accuracy of the translation. That means that translator might add information to the SL text in order to provide deeper understanding in the TL. The opposite to the amplification is the **reduction** which is a process of omitting a further description added to a lexical element in the SL to eliminate redundancy and thus improve the readability of the translated text. The seventh type is the **diffusion**. The diffusion is a process of using more words in the TL than in the SL due to the non-existence of applicable terms or descriptions. It is used to make the SL more understandable to the target audience. The opposite of the diffusion is the **condensation** which is a process of using fewer words in the TL than in the SL due to the existence of an applicable term or description. The last class is the **reordering**. The reordering is a process of rearranging/reordering the original word formation from the SL due to different grammatical rules in the TL. Otherwise, the translated text would not be perceived correctly and thus considered wrong. The goal is to create a more natural expression in the target language.

## 2.3 Terminology

Terminology is the set of unique words belonging to a given profession, a field of study or any social entity. It is a part of applied linguistics studying specialized lexicography. Terminology identifies terms that designate the concepts in the subject field, confirms the usage of the terms in the pertinent reference documents, describes concepts comprehensively or distinguishes the correct usage of terms from the improper usage of terms. (Translation Bureau of Public Works and Government Services Canada, 2001)

Man (2010) further notices that, the terminological element of the text reflects the reality of the professional topic. Terms and terminological expressions are the most distinctive aspect of the professional style. Hrdlička (2010) then adds that, within the given language, terms have both lexical and grammatical meanings. Besides, each term expresses a specific concept which is not bound to any language. Consequently, it can be expressed differently in various languages. A term expressed by one word in the SL might be expressed by a group of words in the TL.

According to Machová (1995), terminology is a linguistic discipline studying, analyzing, describing, and specializing the lexicon. It aims to standardize terms to be used uniformly. A terminological commission discusses any possible changes in terminology at their regular meeting. They keep up with the rapid progress in technological and scientific knowledge.

She also adds that the position of terminology among other linguistic studies is rather unique since it is based on linguistic knowledge as well as non-linguistic disciplines. Knowledge of linguistics mainly consists of morphology, lexicology, and lexicography and is essential for anyone interested in terminology. If linguists deal with a particular area of terminology, they must understand its linguistic structure and meaning while having a basic overview of non-linguistic disciplines. Therefore, the information about terms and its description are present in dictionaries of terminology as well as in general dictionaries.

Moreover, as Poštolková (1983) notices, terminology is an indivisible part of every scientific field. Scientific fields must be exact in their research; therefore, the terminology

must also be adequately precise and unambiguous. Terms have been continuously added to the terminology of a particular field by linguists with knowledge of the given field. If the word from a naturally spoken language becomes a term, it untangles itself from previous meaning relations and enters the network of new relations in the newly acquainted field of study.

Based on previous research, Hrdlička (2010) points out that, the terminological structure and its components must be researched separately, however, it must still be in the system of terms and expressions. The majority of terms came from the formal language, nevertheless, there is a substantial differentiation of the meaning in the professional style. Moreover, terms might even lose their general meaning – they might be marked as meaning-independent.

## 2.4 Lexicology

Lexicology is a linguistic study focusing on the vocabulary of a given language. According to Mitter (2019), it studies and analyses the form, content and functions of particular expressions and their mutual relations. Those words and expressions can be expressed in various ways: a word, phrase, multi-verbalized form, phraseology, and abbreviations. Lexicology also specifies terms, studies the systematic part of the vocabulary, and diversifies the vocabulary. It also studies the composition of vocabulary (nucleus, peripheries, ...) and observes current development tendencies.

Jegede (2020) adds that, typical elements of lexicology in professional texts are long and big words, nominal words, and technical terms. Long and big words are frequently used due to their ability to express the meaning in a more accurate way; however, those words must be in accordance with the given scientific text requirements to avoid ambiguity in words or expressions. Nominal words are used to indicate the formality of the text and to make the text more logical and coherent. Nominal words are often used instead of verbs and adjectives. Technical terms usually carry special meanings different from the meaning we usually face and understand. The use of technical terms makes the information and the content more professional and scientific.

## 2.5 Difference between terminology and lexicology

Based on the information contained in subchapters 2.3 and 2.4, we can see that terminology and lexicology are closely related fields of linguistics; however, they have different aim and scope.

The terminology studies specialized vocabularies that are used in a given field – in the case of this thesis, linguistics, electrical engineering, and informatics. The aim of the terminology is to distinguish how the terms are created and used. On the other hand, the lexicology deals with words and their meanings. It studies how those words were created, their relationship to other words, and how their meanings changed over time. The lexicology also describes and analyses the vocabulary of a language, and its meaning context-wise.

## 2.6 Terms and their translation (general vs. specialized dictionaries)

Translation of a professional text and its terms might seem relatively undiscerning for its seemingly all complications being in the translation of the given terminology, therefore a translator does not have to transfer style characteristics of linguistic means, connotations, and expressivity. Quite contradictory, a translator should possess knowledge of the translated field of study to understand its terminology to be able to handle a lower amount of accessibility and comprehension of this style. (Knittlová, 2010)

Almost every scientific branch has its own unique terminology, therefore, the translator often does not translate terms and, instead, substitutes them. However, if the terminology of the specific scientific branch is specified in both the source and target language, an equivalent term is substituted without any further ado. On the other hand, if the terminology is still being developed and specified, a translator must translate terminology within the word-formation rules of the TL. (Man, 2010)

### 2.6.1 General-purpose dictionary

Britannica defines a general-purpose dictionary as a compilation of other dictionaries with a particular set of users in mind. The general public expects to find there the most



conventionally used terms. Whilst having covered the most common conversational expressions and the most occurring samples from the professional terminologies, the translator should use a dictionary with terminology specific to the given topic for the translation of the professional text.

#### 2.6.2 Specialized dictionary

Each scientific field has its own terminology and has its specialized dictionary. Those dictionaries contain all phrases, expressions, and terms of the given field of study. The general public expects to find there the terminology and lexicology of the given field. However, it does not contain common expressions used in spoken language or other scientific branches, therefore, it is recommended to exploit another dictionary with a specialized dictionary whilst translating.

### 3 Research objectives and methodology

The first part of this chapter will describe the objectives of analysis. Next part will offer a summary of the materials that will be analyzed. The key part will be the subchapter “Aspects of the analysis” which will be divided into four levels: Lexical level, Syntax level, Text level, and Pragmatic level. The final part will provide a description of the procedure that will be used to analyze the texts.

#### 3.1 Objectives and research questions

The aim of the thesis is to analyze four articles that focus on the same topic – the chatbot using Artificial intelligence (AI) Chat GPT. The objectives of the analysis are as follows: to identify any patterns or stylistic features that are typical of the style of science and technology; to evaluate the objectivity and neutrality of the language and identify any instances of bias or subjectivity. Moreover, the analysis will also investigate if the used language is unambiguous and coherent, thus conveys meaning correctly and at the same time will try to find linguistic features and forms appearing in the texts. It will be essential to determine the extent to which the language used in the article reflects the technical and professional nature of the subject matter and possibly identify any linguistic strategies or techniques used to appeal to the reader’s interest and engage their attention. The close observation of the vocabulary in these articles will be made since all four articles focus on the same topic. Lastly, research findings will be examined and repeating patterns or major dissimilarities identified.

Research questions will be then following:

- Does the text show any patterns or stylistic features? If so, do the patterns and features repeat in all four articles?
- Does text show any instances of subjectivity? Is the language neutral?
- What is typical for the used language? Do the texts have the same vocabulary? Do they use terms typical for the language of science and technology?
- What are linguistic similarities and differences among these four articles?

## 3.2 Research materials

The analysis will focus on four articles about AI chatbot Chat GPT that were written by different authors. The articles are taken from well-known news, The Guardian, Business Insider, and BBC Science Focus. The last article is generated by AI and it basically writes the article about itself. All four articles have an overall length of about 2-4 pages. Lastly, the articles were published between March and April 2023 and are available in the Appendices part at the end of the thesis.

The first article was published by The Guardian in March 2023 and is written by 2 authors – Alex Hern and Johana Bhiyan. The authors describe an updated version of chatbot powered GPT-4 released by the owner of Chat GPT company OpenAI. The text mentions new possibilities such as accepting images as inputs and asking questions about pictures. The authors also describe partnership with companies like Duolingo and Stripe so Chat GPT can offer the best experience. The text has 757 words and a total of 4,568 characters.

The second article was published by Business Insider also in March 2023. The author of the article is a journalist Sindhu Sundar. The article discusses the release of ChatGPT chatbot and how it works. In spite of generating impressive responses, the chatbot still lacks human-like thinking. The author also mentions other companies which are trying to develop a competitor to OpenAI's product. Finally, the author discusses the use of the chatbot to solve real-world problems or provide services. However, it still has limitations and can generate incorrect or even nonsensical responses. The article consists of 558 words with the amount of 3,542 characters.

The third article was published by BBC Science Focus in early April 2023 and was written by a journalist Alex Hughes. The author writes about the launch of a new major upgrade for chatbot ChatGPT. After the update, the chatbot can generate texts which are more human-like and can translate language. There is also a new option of 20\$ monthly subscription in return for priority access, faster load times and higher amount of word input level. The article is 1,156 words long and has 6,805 characters.

The last article is an AI experiment. I asked ChatGPT to generate a news article with the following questions: „Write a news article about ChatGPT. Tell me what it does and its major advantages and drawbacks. The length of the article should be 2 A4 pages“ and „Imagine you issue the article in the news. How would you name it?“. The chatbot then generated an article about itself, a large language model trained by OpenAI that can communicate with words inputs in a natural and intuitive way. Then explains how Chat GPT works, describes its advantages and potential drawbacks. The article is 541 words long and has 3,210 characters.

### 3.3 Aspects of the analysis

There are four levels that will be considered: Lexical level, Syntax level, Text level, and Pragmatic level. The **lexical level** will observe the vocabulary and overall language the authors used in their article. Given the fact, all four articles focus on AI (thus on technology), the language should use terms from the style of science and technology. The lexical level also observes nominalization and noun groups. Moreover, the professional text should be unbiased and objective, any example of subjectivity is unsuitable for such type of an article. As has already been mentioned, texts are written in the style of science and technology, therefore, the texts will contain professional terms such as internationalisms, acronyms, descriptive metaphors and adaptations. In addition, the lexical level might also observe figurative language such as metaphors, similes, and idioms. There might also appear some cases of dialect or slang. Overall, the lexical level analysis involves vocabulary, structure, style, and tone of the language used to better understand the meaning and message of the text.

The **syntax level** analyzes how words are combined to create grammatical structures, and how these structures contribute to the overall meaning of the text. It studies various aspects of sentence structure in the text. For example, if the sentence is simple, compound, and complex, what the arrangement of phrases and clauses within sentences are, if the author uses different grammatical structures such as passive voice, impersonal sentence structures, and conditional clauses, and if sentences are cohesive. Therefore, it also analyzes the use of punctuation, such as commas periods, and semicolons, which helps to clarify the

relationship between different parts of a sentence. Alongside the use of punctuation, the use of conjunctions and transitions is observed. These may help connect ideas and create better coherence. To summarize, syntax level examines structure and organization of sentences, phrases, and clauses in order to better convey the meaning of the message and make it more unambiguous.

The **text level** analyzes how individual sentences and paragraphs are combined to create a larger, more complex text, and how these parts work together to make the text coherent. It may examine the way the text is divided, the use of headings and subheadings, and the way in which ideas are introduced, developed, and concluded. Moreover, the use of rhetorical devices such as repetition, contrast, and parallelism is observed. In summary, the text level analyzes the visual level of the text and helps to reveal the overall message and purpose of the text, and how different linguistic elements contribute to this meaning.

Finally, the **pragmatic level** analyzes the social and situational context in which language is used, as well as intention of the speaker or writer and the expectations of the audience or reader. In practice it shows modality and hedging. The use of specific language choices and structures to convey social meaning, such as politeness or formality is examined. Moreover, the use of discourse markers guiding listener's interpretation of the message is analyzed. Overall, the analysis of the pragmatic level examines the way in which language is used in context to achieve specific communicative goals, and how different linguistic elements contribute to the interpretation and understanding of the message.

### 3.4 Procedure

The testing procedure will employ four parts of each text with each text being analyzed in terms of lexical level, syntax level, text level, and pragmatic levels. As it was mentioned earlier, the lexical level will consist of vocabulary and language analyses such as appearance of dialect, internationalisms, archaisms, slang, formal or informal language, specialized terminology, and lexical density or sparsity. The syntax level will analyze the use of grammar, complexity of sentences, discourse functions and passive/active voice structures. The text level will analyze a visual level of the text such as how the text is structured, how ideas are described, repetitions, and overall coherence. Finally, the

pragmatic level will analyze social and situational context, modality, hedging, formality and if the text fulfills communicative goals.

After finishing the analysis of each one of four texts, the results of the analysis will be evaluated and compared. If the text has any specific features or patterns, it will be compared to other texts which might share these features or not. Comparison will also highlight aspects these texts have in common and also things in which they differ. The language will be also a subject of comparison. Whether they show any instances of subjectivity, contain neutral words, use same or different vocabulary, and contain terminology for particular field, or not. It will be basically a comparison of analyses of all four texts and finding out why these phenomena in the analyses happen.

The most interesting phenomena which will be provided with an example from the texts, will be translated into Czech language. The translation will be then analyzed and the processes of translation will be described.

### 3.5 Summary

The third chapter contains information about the main objectives of the practical part and how the analysis will be carried out. To summarize, the lexical level analyzes language and vocabulary of the text, the syntax level analyzes how the sentences are structured, the text level analyzes visual level of the text, and, lastly, the pragmatic level analyzes context. These four analyses will then be compared with each other.

## 4 Analysis

The analysis should consist of the Lexical level, Syntax level, Text level, and Pragmatic level. The final part of the project is a conclusion where all findings should be discussed. Each lexical analysis will contain a table where *Word class X* stands for the most frequent word classes, *Word X* stands for the most frequent words in the given word class, and, lastly, *n* stands for the number of appearances of the given word. Tables with the most frequent terms in word classes will not contain names.

Some examples will be chosen for translation. These examples will be marked by  $(n)$  before the sentence, where  $n$  means the number of examples.

### 4.1 Text 1

#### 4.1.1 Lexical level

Given that the text is focused on information technology, some repeatedly used words might create a perception of lexical sparsity. The name of the product is the most occurring noun in the article – *GPT-4*, which appeared fifteen times in the text. The second most occurring noun is *users*, which appeared nine times. The name of the company *OpenAI* was mentioned eight times alongside the word *model*. Among the most used nouns are also words *version* and *system*, which were mentioned six times and *images* and *language* with five mentions.

The most frequent words		n
Word 1	GPT-4	15
Word 2	Users	9
Word 3	OpenAI, Model	8

Table 1 Text 1 - The most frequent terms

Because the text is focused on one of the latest technologies (specifically, its most recent update), the vocabulary is also specific. It is marked as a *language model* which is why these two words are highly occurring. The new update (or *version*) of the Chatbot (reason for *language*) allowed *users* to have *images* as an input or output. As I tried to emphasize

in the previous sentences, there is a limited vocabulary within the topic, therefore, it is vocabulary-wise understandable and predictable.

Considering other word classes, out of adjectives, the most occurring word is *new* with seven occurrences and *latest* with three occurrences. Given that it is the latest technology, as mentioned earlier, the usage of these adjectives is also understandable and predictable. As shown in the table below, the most frequent terms in the register of science and technology are nouns *users*, *model*, *version*, and *system*. The most frequent technological adjectives are *capable*, *artificial*, and *creative*. Lastly, the most frequent verbs were *submit*, *create*, and *improve*.

	Word class 1		Word class 2		Word class 3	
	Nouns	n	Adjectives	n	Verbs	n
Word 1	Users	9	Capable	2	Submit	2
Word 2	Model	8	Artificial	2	Create	2
Word 3	Version, System	6	Creative	2	Improve	2

Table 2 Text 1 - The most frequent terms in word classes

The article includes technological terminology such as the abbreviation *AI* or in its complete form, *Artificial intelligence*, *model*, *image*, *input*, *output*, *prompt*, *data*, *app*, *system*, *scammer*, *support forum*, *demo*, *tested*, *prompt*, and many other IT terms. Many of these terms are also internationalisms, such as *AI*, *input/output*, *data*, *system*, *scammer*, and, for example, *demo*. Moreover, the article shows no glimpse of dialect or sociolect, and it is written in formal Standard English. However, there are cases of slang such as *app* (application), *demo* (trial version), and, for example, *prompt*.

Internationalisms have usually same form in Czech language as in the source language, if the form is the same, the translation procedure is borrowing. For example, *AI* is *AI* even in the Czech language – abbreviation *UI* could be used (for *Umělá Inteligence*), however, it is not used at all in that abbreviated form. Other examples are *data*, which is *data* even in Czech language, *input/output*, which might be substituted by *vstup/výstup* but *input/output* is still often used in Czech language, or *scammer*, which could be substituted for *podvodník*, however, *scammer* is used specifically for online frauds even in the Czech language. Lastly, Czechs usually use special type of borrowing in which the original term



is almost identical, however, Czech equivalent is slightly modified, such as *system*, which is *systém* in Czech, or *forum*, which is *fórum*.

Translations of professional terminology are usually set in the language or they are internationalisms, therefore, it is either calque or borrowing. Examples of calque are *artificial intelligence* – *umělá inteligence*, *image* – *obrázek*, *support* – *podpora*, or *app* – *aplikace* (or as a slang *appka*). Borrowing examples are similar to borrowings in previous paragraph, such as *model* – *model*, *data* – *data*, or *scammer* – *scammer*. Only term, which is slightly different in Czech is *demo*, which is *demoverze* in Czech. The translation adds information; therefore, it is equivalence.

#### 4.1.2 Syntax level

The most frequent word class was nouns. Far less used (but still second frequent) were verbs and prepositions. The language or technology is highly terminological; therefore, it is unsurprising that nouns are the most frequent word class. Adjectives were fourth, and adverbs fifth in terms of frequency. These word classes are regularly used in the whole text.

The sentences in the article are rather complex. This is standard practice in the register of technology journalism. Sentences are usually complex or compound-complex. Complex sentences often include citations from the announcement or an adjective-dependent clause. Despite being rather complex, sentences are cohesive and understandable. To prove the complexity of sentences, the period was twenty-four times in the text, whereas the comma was more frequent with thirty-nine cases.

The article includes some instances of passive voice structures; however, it is not very frequent. Passive voice structures are usually used with inanimate objects in the technological register. The following examples show passive voice structures and their translations:

- (1) *...of the newest version of the system, which is not yet publicly available and **only being tested** by...*  
*...nejnovější verze systému, která zatím není veřejně dostupná a zatím **se pouze testuje**...*

The translation is an equivalent of passive voice in Czech language. Therefore, the phrase *only being tested* can be directly translated as *zatím se pouze testuje*. The literal translation would be *pouze je testovaný*, however, I added *zatím* to specify that the current state of the process is only temporary. The translation is an equivalent according to translations procedures by Vinay and Darbelnet (see 2.1).

- (2) *...GPT-4 **was asked** to describe...*  
*...GPT-4 **byl požádán**, aby popsal...*

This translation is a calque according to Vinay and Darbelnet (1958), therefore, it is the literal translation of the source text.

#### 4.1.3 Text level

From a visual point of view, the text could be more diverse. The article is divided into eleven paragraphs ranging from three to nine rows in terms of length. Each paragraph has the same font type and font size. It has a title written in a larger font size and seemingly in bold letters. There is a short description of the content below the title written in bold letters. Before the article begins, we can also see both authors' names and locations with the date of publishing.

Moreover, the article contains one image and one video. Both have footers which include short descriptions and authors. The length of the article is 754 words and a total of 4,581 characters including spaces, spread out over two A4 pages.

The text has a title with a short text below describing the content of the article. However, there are no sub-headings which would in my opinion help the overall coherence. There are no topic shifts in the text because the topic is only one. The authors describe what happened during the OpenAI announcement in chronological order.

There are only a few cohesive devices in the text. The authors use the word *it* quite often and refer to GPT-4 with that. Given that authors write about a current announcement, they refer to what was said by the creators. Using cohesive devices is expected in the technological register, however, this text is a short news article without any chapters or subchapters. Thus, it would be uncommon for such a short article to have a lot of cohesive devices.

#### 4.1.4 Pragmatic level

The article is meant to be informative, and the impact on a reader is informative. In my opinion, the main intention of the authors was to summarize the event so people do not have to watch the whole conference. According to Jakobson's language functions (Zmrzlá, 2022), the text has a referential function and might have a conative function (they do not want people to watch the conference and read a shorter article instead). Another function worth mentioning is metalingual, since the chatbot is a language model and by talking about its word generation the authors also talk about the language itself at the same time.

There were only a few instances of modality because authors write about a real event and refer to situations that already happened. It is meant to be objective. Following examples illustrate the use of modality:

(3) ...***that can mean that*** *it makes up information when it doesn't know the exact answer...*

...***toto může znamenat, že*** *si vymýšlí informace, když nezná přesnou odpověď...*

(4) ...*the system is particularly good....*

...*system je obzvláště dobrý...*

The first case shows us authors worry about the credibility of AI generated answers. On the other hand, the second case points out that the system is outstanding in something or above average.

Examples 3 and 4 are the examples of modality, therefore, modality must be included in the Czech translation. The modality phrase *can mean that* shows the level of uncertainty of the idea is translated as *může znamenat, že*. The target language has equivalents for the

phrase; therefore, the translation is literal. According to Vinay and Darbelnet translation procedures (1995), the translation is calque. Similarly, the other example *particularly* is also an example of modality and emphasizes the quality of the chatbot. Czech language offers an equivalent *obzvláště*, therefore, the translation is calque again.

Semantically, the text is straightforward and unambiguous because it is written in the style of science and technology. The possible use of figurative language might be that processes in the application are processed spontaneously, although the application always needs an input to provide an output. Another case might be that AI thinks and creates, whereas it only processes a huge amount of data to provide an output. This might be an example of figurative language because the author considers inanimate program animate.

#### 4.1.5 Conclusion

In the first part, we introduced and repeated information about the text. Secondly, in the lexical part, the vocabulary mainly consisted of terms from the style of science and technology. The most mentioned words in the text (apart from the articles and otherwise commonly used words) were *GPT-4*, *users*, *model*, and *OpenAI*. The text looks lexically sparse because the terms are repeatedly appearing in there, however, it is lexically dense regarding the usage of less professional terminology. Thirdly, in the syntax part, we analyzed that the most occurring word class are nouns which was unsurprising given that the authors used vocabulary from technological terminology register. Sentences were mostly complex, with a substantial number of citations and adjective-dependent clauses. The article was descriptive and objective, with few passive voice structures. In the text part, we found out that the text has only a few instances of cohesive devices. It is a formal text, informative and has two functions according to Jakobson's – referential and conative. From a visual point of view, the text uses one type of font with the same size, the only difference is the title and short information below. It is divided into visually similar paragraphs. Finally, in the pragmatic part, we found only two instances of modality where the in first case authors worry about credibility and in the second case highlight a positive feature of the AI tool. Moreover, figurative language is not often used in the text. The only

case that might be considered as figurative language is when authors consider an inanimate PC application to be animate.

Overall, the text was similar to what I expected it to be. I supposed that the text would be formal with citations, informative, structure-wise not very divided, and using vocabulary from the technology register. The only thing I expected that did not occur was the appearance of vocabulary from the business register.

The authors achieved their intention to summarize what was happening during the OpenAI announcement. They used a picture and a video to help to understand the topic and tried to explain the matter even for a non-professional audience. The sender achieved communicative goals.

## 4.2 Text 2

### 4.2.1 Lexical level

The second text also uses vocabulary for the language of science and technology. The main focus here is AI and the chatbot ChatGPT; however, the article does not mention the newest update GPT-4, because it was not issued at that time. The most frequently appearing word class were nouns, which is a typical thing for the language of technology. Except for nouns, the most used are verbs, prepositions, and adjectives. As shown in the table below, the most frequently used terms are *chat*, *ChatGPT*, *bot*, *language*, *users*, *tools*, and *AI*.

The most frequent words		n
Word 1	Chat, ChatGPT	7
Word 2	Bot, Language	6
Word 3	Users, Tools, AI	4

Table 3 Text 2 - The most frequent terms

After a more thorough analysis of the vocabulary, the most appearing nouns are *ChatGPT* and *chat*, where both have been used seven times, *bot* and *language* with six appearances, where *bot* is usually used along with *chat* (*chat bot*), and *users*, *tools*, and *AI* were used four times. The second most occurring word class were verbs. As shown in the table below, examples of terminological verbs that were mentioned are *experiment*, *simulate*,

and *evoke*, and the adjective that was the most frequent was *large*. *Large* was the only adjective mentioned more than two times, which means, the text is lexically diverse. Examples of adjectives from the professional style are *viral*, *instant*, and *quantum*. Professional verbs and adjectives were not repeated very often; therefore, the article is not professional in terms of language of science and technology but popular scientific.

	Word class 1		Word class 2		Word class 3	
	Nouns	n	Verbs	n	Adjectives	n
Word 1	Chat	7	Experiment	1	Viral	1
Word 2	Bot, Language	6	Simulate	1	Instant	1
Word 3	Users, Tools, AI	4	Evoke	1	Quantum	1

Table 4 Text 2 - The most frequent terms in word classes

Since the article uses language within the register of science and technology, there are some internationalisms, vocabulary specific to the register, and slang. Internationalisms are, for example: *chatbot*, *chat*, *internet*, and *AI*. Examples of vocabulary specific to the register of science and technology might be: *command*, *chat*, *search engine*, *web*, *browse*, *data set*, *AI*, *tool*, *user*, *plug-in*, *system*, *model*, or *interface*. Lastly, examples of slang terms are: *IRC chat room*, *bot*, or *prompt*.

Interesting examples of translations which do not appear in the subchapter 4.1 are internationalisms *chat* – *chat*, *chatovat* (depending on word class – noun, verb), *chatbot* – *chatbot*, which could be translated as *chatovací robot*, but the term is not used, and *internet* – *internet*. Internationalisms are usually translated by borrowing. Examples of professional vocabulary translations are *command* – *příkaz*, *search engine* – *vyhledávač*, *browse* – *prohlízet*, *data set* – *datový soubor*, *IRC chat room* – *IRC chatovací místnost*, or *plug-in* – *plug-in*, *doplněk*. Most of these examples are translated as a calque, however, *plug-in* in Czech is borrowing and *chatovací* is a modulation because *chat* is noun in Czech and it needs to be modified to adjective by using a suffix. Lastly, the translation of the slang word *prompt* is *řádek*, which is also a calque, or it could be translated as *příkazový řádek*, which would be amplification.

Overall, the article includes terms specific to the register of science and technology. Therefore, these terms are repeated in the text. However, apart from the specific

terminology, the author uses lexically dense vocabulary. Part of the article are citations from oral speeches of the specialists in the given field, which are more lexically sparse than the text.

#### 4.2.2 Syntax level

The syntax is generally straightforward and follows standard grammatical structures. As it was mentioned earlier, the most common word classes are nouns, verbs, and adjectives.

Sentences have mostly complex structures, usually to convey more detailed information or present multiple ideas in a sentence. The following example shows a complex structure:

- (5) *"There's a saying that an infinite number of monkeys will eventually give you Shakespeare," said Matthew Sag, a law professor at Emory University **who studies copyright implications** for training and using large language models like ChatGPT.*

*"Jedno přísloví říká, že nekonečný počet opic vám nakonec napíše Shakespeara," řekl Matthew Sag, profesor práva na Emory University, **který se zabývá důsledky autorských práv** pro trénování a používání velkých jazykových modelů, jako je ChatGPT.*

Another characteristic is visible in the example, and that is the number of citations. The author uses quoted speech to emphasize the opinions and statements of the specialists who shared their thoughts on the given matter. The quotations follow standard rules of punctuation.

Example 5 is the example of complex sentence, which is formed by possessive sentence *who studies copyright implications* and the Czech translation is *který se zabývá důsledky autorských práv*. Where *who – který* is a modulation to make better sense of the sentence. The literal translation of *který* is *which*. The last thing worth pointing out is word *copyright*, which is translated as *autorské práva*, that translation process is amplification.

Given the fact that AI is a very controversial topic nowadays, the author also uses contrastive and comparative phrases. The use of contrastive sentences and phrases usually reflects some drawbacks after highlighting a benefit. The following example shows the contrastive sentence:

- (6) *"There's a large number of monkeys here, giving you things that are impressive — but there is intrinsically a difference between the way that humans produce language, and the way that large language models do it," he said.*

*"Je tu spousta opic, které vám poskytují pozoruhodné věci, ale existuje vnitřní rozdíl mezi způsobem, kterým lidé vytváří jazyk, a způsobem, kterým to dělají velké jazykové modely" řekl.*

The contrast is maintained, and the meaning of the sentence is still the same. The source language has an equivalent for translating a contrastive phrase, therefore, the translation procedure is calque. The literal translation of *he said* is *on řekl*, however, Czech language use only *řekl*, which would be literally translated as *said*. This procedure is called condensation according to Malone (1988). Another example worth pointing out is the use of dash, which is used in English texts but in Czech texts only rarely. Lastly, giving you things is a transitive phrase, which is considered archaic in Czech, therefore, I used a possessive phrase instead. The opposite thing happens with the sentence that are impressive which is translated into a single word *pozoruhodné*, that is a process of condensation.

Moreover, the author uses parenthetical phrases, which provide additional information or clarify unambiguous phrases in the sentence. The example below illustrates the use of parenthetical phrases:

- (7) *To older millennials who grew up with IRC chat rooms — **a text instant message system** — the personal tone of conversations with an AI bot can evoke the experience of chatting online.*

*Starším mileniálům, kteří vyrostli na chatovacích místnostech IRC, **systemu okamžitých textových zpráv**, může osobní tón konverzace s umělou inteligencí nabudit zážitek z online chatování.*

The main reason for this example is the parenthetical phrase *a text instant message system*, which is translated as *systemu okamžitých textových zpráv*. The main difference is the use of dashes as I mentioned earlier in the subchapter. Czech usually uses commas instead of dashes. Otherwise, the sentence is translated as a calque with the use of reordering to make the sentence more comprehensible.



The article also contains passive voice structures in some instances. The passive voice is usually used to emphasize the action rather than the subject. The following examples illustrate the use of passive voice:

(8) ...*ChatGPT was trained*...  
...*ChatGPT byl trénován*...

(9) ...*Chat bots like GPT are powered* by large amounts of data...  
...*Chatboti jako GPT jsou poháněny* velkými množstvími dat...

The use of passive voice structures has an equivalent translation in Czech, therefore, *was trained* is translated as *byl trénován* and *are powered* as *jsou poháněny*. Both cases are translation processes of calque.

#### 4.2.3 Text level

From a visual perspective, the text looks rather bland. There is only one main heading without any subheadings or any other subdividing aspect. The article is structured into sixteen short paragraphs ranging from three rows to seven rows. The article consists of 3,573 characters forming 355 words and spreading over two A4 pages. The title is written in a larger font and in bold letters, and all paragraphs have the same font and formatting. There are also underlined parts containing a link to different articles focusing on the mentioned matter in the article and quotes addressing speeches of various professionals. The article contains one stock image with the logo of OpenAI and ChatGPT. There is also the name of the author, with the issue date, time, and time zone above the picture. From a visual point of view, the article can be more diverse.

The text also contains several cohesive devices, such as: *it*, which refers primarily to *ChatGPT*, *they*, or *them*. Other instances of cohesive devices might be contrastive conjunctions referring to previously said sentences or phrases, such as *but* and *though*. Lastly, citations can be considered as cohesive devices. Those quotations occur frequently in the text, and the author uses them to refer to statements and opinions said by specialists in the field of AI and technology.

#### 4.2.4 Pragmatic level

According to Jakobson's language functions (Zmrzlá, 2022), the article has a referential function. Since the text is about AI language models, another function is metalingual because language is used to describe language. Moreover, the text is informative, and the author summarizes the knowledge about the given topic and forms it in a way that conveys general information about the matter in the most plausible and comprehensible way for readers.

The author also employs modality either to show their own opinion or worries or to make the statements less/more expressive. The following examples illustrate the use of modality:

- (10) ...*a full AI takeover **isn't exactly imminent**, according to experts...*  
...*podle odborníků úplné převzetí umělé inteligence **přesně nehrozí**...*
- (11) ...*other researchers **seem to be taking** more measured approaches...*  
...***zdá se**, že jiní výzkumníci volí rozvážnější přístupy...*
- (12) ... ***it seems a matter of time** before...*  
...***zdá se**, že je pouze otázkou času, než...*

In examples 10 and 11, the author tries to make the text less expressive and make the statement more uncertain. Example 12 shows the author's assumption of what might happen in the future. As can be seen in the examples, modality is used to change the expressiveness of the statement or, for example, to show the subjectivity of the author.

The translation of example 10 reflects authors intention to lower the general worries about the takeover of AI. As already mentioned in the previous paragraph, in example 11 the author tries to make the statement less certain, which should be incorporated in the Czech translation. There is no literal equivalent for *seem to be* in the Czech language; therefore, I used a modulation *zdá se, že*, which would be literally translated as *apparently*. However, the overall meaning and the intention to make the statement less certain remains in the translation. The same thing happens in example 12.

The text is coherent and unambiguous, which is a typical feature of the language of science and technology. Consequently, there are not a lot of occurrences of figurative language in

the text. The author considers an inanimate chatbot animate, which might be an example of metonymy.

#### 4.2.5 Conclusion

In conclusion, the analysis shows us that in terms of vocabulary, the article employs words within the language of science and technology, therefore, the author uses their vocabulary accordingly to the register. The most occurring word classes are nouns and, subsequently, verbs and adjectives. Repetition of terms is a feature of the language of science and technology due to the limitations of synonyms to professional terminology. Apart from the terminology of science and technology, the text is lexically dense. The syntax part of the analysis shows us that the sentences in the text are mostly complex, and the author uses a substantial number of citations. There are also some instances of the passive voice, however, it is not that frequent. From the visual point of view, the text looks very plain and not eye-catching. Lastly, pragmatically, the texts have a referential function, thus, their primary purpose is to inform. The article also contains some instances of modality showing the author's opinion or their attempt to change the expressiveness of the phrase or sentence.

### 4.3 Text 3

#### 4.3.1 Lexical level

The article published by BBC Science Focus deals with the topic of ChatGPT and its newest update GPT-4. The language used in the text falls within the register of science and technology, therefore, it is logical to presuppose that the article will contain a significant number of terminological vocabularies. The word classes that appear the most are again nouns, which is a typical feature of the language of science and technology. Subsequently, the other most used word classes were verbs, prepositions, auxiliary verbs, and adjectives. Counting up the most occurring terms in the text, the most frequent is the name of the product *ChatGPT* with twenty-six occurrences, the name of the update *GPT-4*, which appeared twelve times, and *AI*, which also appeared twelve times.

The most frequent words		n
Word 1	ChatGPT	26
Word 2	GPT-4	12
Word 3	AI	12

Table 5 Text 3 - The most frequent words

As in the previous paragraph, the most common word classes in this text are nouns, verbs, prepositions, and adjectives. Among nouns, the most used words were *ChatGPT* with twenty-six occurrences and *GPT-4* with twelve. Apart from the names, the most used nouns were *AI*, which was there twelve times, *language* appearing eight times, and *version* with seven occurrences. The most frequently used verbs were *create*, which appeared five times, *generate* with two appearances, and *demonstrate*, which also occurred two times. Lastly, since the article is concerned about the newest version of ChatGPT, there is no surprise that the most frequent adjective was *new*, which appeared six times in the text. The examples of frequently used professional adjectives were *worded* with four occurrences, *artificial*, which appeared two times, however, the abbreviation *AI* appeared twelve times and the word *artificial* is a part of it, and *powerful*, which also appeared two times.

	Word class 1		Word class 2		Word class 3	
	Nouns	n	Verbs	n	Adjectives	n
Word 1	AI	12	Create	5	Worded	4
Word 2	Language	8	Generate	2	Artificial	2
Word 3	Version	7	Demonstrate	2	Powerful	2

Table 6 Text 3 - The most frequent terms in word classes

Given that the article utilizes the language of science and technology, the author uses professional terminology in that field. Therefore, there are instances of internationalisms, technological terms, and slang. The following sentences will provide examples of each of these phenomena. Internationalism appearing in the text are, for example: *AI, model, chat, ban, platform, or input*; technological terms are: *software, transformer, application, program, script, or ban*; lastly, slang terms are: *bot, prompt, hallucination, or pro version*.

The examples which were not mentioned in earlier subchapters are, for example, *ban, software, platform, or pro version*. Ban is a term used often in it and the literal translation would be *zákaz/zakázka*, however, in Czech IT slang is more common *ban/dát ban*, which would be borrowing for ban and amplification for *dát ban*, because the literal translation would be *to give a ban*. Software is software in Czech, platform is platforma, and slang term pro version could be translated as *profi verze*. All these terms are translated as calque.

The author repeats the main terms in their article, however, apart from the professional terminology, the text is lexically dense with few repetitions. An example of lexical density might be the fact that adjectives are one of the most used word classes in the text, however, three of the most used adjectives have six, three, and two appearances. It follows that the author used diverse vocabulary not to repeat the same terms.

#### 4.3.2 Syntax level

The syntax of this article is typical for the given register and follows standard grammatical structures. As mentioned in the lexical part, the most frequent word classes are nouns, verbs, prepositions, and adjectives.

The author uses rather complex sentences in the text to provide extra details for their sentences or to convey more ideas in a single sentence. The example is: *They are capable of generating human-like text and have a wide range of applications, including language translation, language modelling, and generating text for applications such as chatbots.* They used a complex structure to provide more information about the main sentence. An example of a complex sentence used to convey more ideas in a single sentence might be: *Previously, ChatGPT could become confused, offering up a nonsensical answer to you question, or even inputting stereotypes or false information.* The author not only conveys more ideas but also graduates the sentence at the end.

Another feature of the article is the number of contrastive phrases since the topic is considered to be controversial. There are many specialists who see AI as a benefit, and many specialists say it is a threat. The following sentences provide an example of contrastive sentences: *It's a little bit boastful, but completely accurate and arguably very well written* or *Currently, ChatGPT remains free-to-use software. However, OpenAI has now announced ChatGPT Pro – a pay to use version with added benefits.* The author tries to make statements in the examples less expressive by providing a contrastive statement in the other sentence.

(13) *It's a little bit boastful, **but completely accurate** and arguably very well written.  
Rád se trochu chlubí, **ale je zcela přesný** a nepochybně velmi dobře napsaný*

(14) *Currently, ChatGPT remains free-to-use software. **However**, OpenAI has now announced ChatGPT Pro – a pay to use version with added benefits.  
V současné době zůstává ChatGPT bezplatným programem. **Nicméně** společnost OpenAI nyní oznámila ChatGPT Pro, placenou verzi s dalšími výhodami.*

Both these sentences are contrastive. The first example has contrastive phrase *but completely accurate*, which is translated as *ale je zcela přesný*, that translation procedure is a calque. Contrast in the second sentence is showed by the word *however*, which has a Czech equivalent *nicméně*. The literal translation of *nicméně* is *nevertheless*.

The article also contains parenthetical phrases, which might present some additional information or clarify unambiguous phrases in the sentence. The following examples show parenthetical phrases in the article:

(15) *This would include OpenAI's work on GPT-5 – the next version of technology ChatGPT will eventually run on.*

*Mohlo by to zahrnovat práci OpenAI na GPT-5, další verze na které poběží ChatGPT*

(16) *Its most common use so far is creating ChatGPT – a highly capable chatbot.*

*Zatím nejčastější využití je vytvoření ChatGPT, vysoce schopného chatbota*

Parenthetical phrases in the examples are translated as calque, the difference is that Czech language use commas instead of dashes to mark the parenthetical sentences.

The article uses predominantly active voice structures; however, the author also uses passive voice several times. They use it, for example, when discussing actions performed by AI or describing some concerns. The following examples illustrate the use of passive voice:

(17) *ChatGPT was announced*

*ChatGPT byl oznámen*

(18) *it has been banned in select schools*

*bylo to zakázáno ve vybraných školách*

(19) *called out by governments*

*byli vyzváni vládami*

(20) *utilised by major companies like Microsoft*

*využíváné velkými společnostmi, jako je Microsoft*

Czech language use passive voice structures to translate passive voice structures from English. Therefore, Czech has an equivalent for it and the translation process is calque. The translations of provided examples are *was announced* – *byl oznámen*, *has been banned* – *bylo zakázáno*, *called out* – *byli vyzváni*, and *utilised* – *využíváné*. The difference is in example 18. The use of the tense is different, where in English *has been banned* is in present perfect and Czech translation *bylo zakázáno* is in past simple. In this case, the translation procedure is modulation.

### 4.3.3 Text level

Visually, the text looks diverse. The author uses various subheadings and also a short information about the article just below the title. Below the title, there is also the name of the author, issue date and time. The article is extensive; however, I chose only a part of the article for the analysis, otherwise, the article would be too long. The part that was chosen for the analysis is divided into forty-one paragraphs, which have two to five rows. The paragraphs are primarily concise, and sometimes they consist only of one sentence. There are 1,164 words in the article and a total number of 6,906 characters spread out over four A4 pages.

The article has the main title written in bold letters, but the text is divided by headings and subheadings. There are six different headings and three subheadings (only within the part I have chosen for the analysis, otherwise, the article is much longer and has more headings and subheadings). Moreover, there are hyperlinks to specific topics within the text that are underlined and blue. There are also three pictures. Two seem to be regular stock photos, and the third one is a screenshot of the chatbot ChatGPT. There is a title below the first picture, and below the title is the author and the date of publishing. In addition, the article has one bulleted list with four types of information. From a visual point of view, the text is diverse, cohesive, and easy to focus on.

Lastly, the article contains several instances of cohesive devices. Firstly, there are pronouns such as *it*, *this*, and *they*, which are usually referring to either ChatGPT and OpenAI or the creators. The use of these pronouns helps to maintain the continuity of the text. Secondly, the author repeats main terms and phrases such as *ChatGPT*, *GPT-3*, and *GPT-4* to establish a connection between related information. Lastly, there are also some instances of synonymy where, for example, the author uses *banned* and *prohibited* to avoid the repetition of a single word. Fourthly, the article contains transitional adverbs and conjunctions such as *however*, *also*, or *while* to indicate shifts in information, introduce contrasting or additional points, and establish logical connections between ideas.



#### 4.3.4 Pragmatic level

According to Jakobson's division of language functions (Zmrzlá, 2022), the text has a referential function, and the other functions could be conative because there are some persuasive elements, and metalingual because the author talks about a language model. The text has a primarily referential function because the author intends the text to be informative and tries to sum up the relevant knowledge about the chatbot. To provide reasoning for the conative function, the author highlights the capabilities and achievements of ChatGPT, such as passing legal exams, writing articles, or providing codes to create specific websites. In addition, the article contains rhetorical questions with the intention to engage the reader, for example: *And is this really the future of AI?* or *So what is ChatGPT?*.

The author tries to balance enthusiasm and realism using modality in the article. The author might also want to change the expressiveness or show their uncertainty and worries. The examples of modality in the article are *...artificial intelligence research could be forced to slow down...*, where the author shows his assumption of the future of AI, *Not much has been revealed about this version...*, in which the author provides their opinion about the amount of revealed information, or *It's a little bit boastful, but completely accurate...*, which shows two instances of opposite modality – by *a little bit* the author tries to make the following phrase less expressive and, on the other hand, by *completely* the author makes the following phrase more expressive and confident.

The following sentence shows the translation of one of the examples:

- (21) *...artificial intelligence research could be forced to slow down...*  
*...výzkum umělé inteligence by mohl být nuceně zpomalen...*

The example shows author's assumption about the future situation, which should be included in the Czech translation. The Czech tried to keep the meaning of *be forced to* as *nuceně*, which has the similar meaning while being a different word class. The main modality phrase is *could be* has a direct equivalent in the Czech language, therefore, it is translated as *by mohl být*, which is a calque according to Vinay and Darbelnet (1995).

Finally, since the topic is in the language of science and technology, the text is predominantly unambiguous, therefore, without nearly any instances of figurative language. The only instances with figurative language are when the author considers an inanimate AI program animate or when the author talks about *hallucinations*, which is a figurative term describing previous mistakes made by ChatGPT.

#### 4.3.5 Conclusion

To conclude the previous subchapters, the lexical level of the analysis shows that the most frequent word classes are nouns, which is a typical feature for the language of science and technology, verbs, prepositions, and adjectives. The author uses vocabulary from the register of science and technology, precisely, the register of artificial intelligence. There are also some instances of internationalisms and slang. Apart from repeated terms, which do not have other substitutes, the text is rather lexically diverse. From the syntax level of the analysis, it is apparent that the sentences in there are primarily complex; there are some cases of contrastive sentences and phrases; the author sometimes uses parenthetical sentences to provide additional information; and the article contains mostly active voice structures with some instances of passive voice. The third level of the analysis is the text level, and it shows that the article is visually diverse, using several different headings and subheadings, pictures or one bulleted list. To achieve well cohesiveness of the article, the author uses several cohesive devices. Finally, the text has a referential function with the intention of being informative. There are also some instances of conative and metalingual functions. Moreover, the author uses modality to change expressiveness or to show some amount of subjectivity.

#### 4.4 Text 4

The last article is generated by the AI chatbot ChatGPT. The chatbot was asked these questions: *Write a news article about Chat GPT. Tell me what it does and its major advantages and drawbacks. The length of the article should be 2 A4 pages and Imagine you issue the article in the news. How would you name it?.* The chatbot then generated a text about itself and created a title. In my opinion, providing an example generated by AI shows how far it has gone and what it is capable of. It would be interesting to see how it

stands out against articles written by journalists. The generated text is available in the appendices, and it also shows the date when the text was generated. The text was generated on the currently free version of GPT-3 and not on the more advanced GPT-4.

#### 4.4.1 Lexical level

The AI inspired itself by other articles published by various journalists and followed the same vocabulary of the language of science and technology. The structure of the vocabulary follows the typical structure as well. The most frequently used word class were nouns and in a relatively higher amount compared to other word classes. The second most used word class were prepositions and, subsequently, verbs and adjectives.

The most frequent words		n
Word 1	ChatGPT	18
Word 2	Use	9
Word 3	Information	7

Table 7 Text 4 - The most frequent terms

As it was mentioned in the previous paragraph, the most frequent word class were nouns. The most frequently used professional nouns were *information*, which appeared seven times in the text, *data* and *responses* with six occurrences, and with five appearances were *language* and *model*. The most common professional verb in the article was *provide*, which appeared seven times. The examples of other professional verbs are *generate* and *explore*, which both appeared only once. Lastly, the most frequent adjectives were *accurate*, which appeared four times, *natural*, which occurred three times, and *artificial* with two occurrences.

	Word class 1		Word class 2		Word class 3	
	Nouns	n	Verbs	n	Adjectives	n
Word 1	Information	7	Provide	7	Accurate	4
Word 2	Data, Responses	6	Generate	1	Natural	3
Word 3	Language, Model	5	Explore	1	Artificial	2

Table 8 Text 4 - The most frequent terms in word classes

Given the fact that the chatbot writes about itself, it inspires itself by other articles regarding a similar matter, therefore, it should use the appropriate vocabulary of science and technology. That language should contain internationalisms, technical vocabulary, and

slang. The examples of internationalisms contained in the article are: *model*, *AI*, *data*, *input*, *potential*, *symptom*, or *game-changer*. Technical vocabulary can be: *AI*, *access*, *data*, *input*, *response*, *platform*, *user*, or *transformer-based neural network*. Lastly, examples of slang language could be: *prompt*, *training data*, or *language model*.

The translations of some interesting examples are *borrowings*, *potential* – *potenciál*, *symptom* – *symptom*, *game changer* – *game changer*, and calques, *access* – *přístup*, *response* – *reakce*, or *transformer-based neural network* – *neuronová síť založená na transformátorech*.

The generated article has not only a lot of repeating non-professional words, which are usually substituted by synonyms to avoid repetition. The article is not lexically dense as it could be, because non-professional words have relatively high amount of repetition given the relatively small size of the article.

#### 4.4.2 Syntax level

The AI tries to replicate the structure of other articles in the given field. As it was mentioned earlier, the most frequent word classes are nouns, prepositions, adjectives, and verbs. AI used a relatively higher number of nouns and prepositions than it is usually used.

The sentences the chatbot generated are rather complex. The examples of complex sentences are: *As a large language model trained by OpenAI, ChatGPT has the ability to converse with humans in a way that feels natural and intuitive*, in which the first sentence specifies or provides additional information to *ChatGPT*, *ChatGPT is based on the GPT-3.5 architecture, which is a deep learning model that uses a transformer-based neural network to process and understand natural language input*, where *GPT-3.5 architecture* is further specified in the following sentences, or *Because the platform has been trained on such a large corpus of data, it can quickly process large amounts of information and provide well-informed responses*, in which the first sentence is a reason why the second sentence happens.

The contrastive sentences are not so frequent, however, there are some instances in which these sentences occur. The examples are: *While the platform was designed to provide valuable information to users, it could also be used to spread misinformation or to manipulate public opinion* or *However, as with any technology, it is important to consider the potential drawbacks and to use it responsibly*, where both instances try to provide a counterargument to the previous statement or to indicate some deviations.

The following example was chosen for the translation:

- (22) ***However, as with any technology, it is important to consider the potential drawbacks and to use it responsibly.***  
***Ale stejně jako u jakékoliv technologie, je nutné zvážit možné nevýhody a používat to zodpovědně.***

The translation must keep the contrast of the source text. *Ale* is an equivalent of *however*, where both words are used to create a contrast. Moreover, there is a parenthetical phrase within the sentence *as with any technology*, which is translated as *stejně jako u jakékoliv technologie*. The original sentence should still make sense even if the parenthetical sentence is removed, which happens in both the original and the translated version. Lastly, an interesting phenomenon in translations from English to Czech is that English needs to have an apparent subject, such as, *it is*. That can be translated, as can be seen in the whole translation, as *je* which would be literally translated as *is*. The Czech version does not need to have the subject, in our case pronoun *it*, which is a condensation according to Malone (1988).

There are only a few instances of parenthetical phrases which might provide a solution to ambiguity or present some additional information. The examples of parenthetical phrases are, for example: *As a large language model trained by OpenAI, ChatGPT has the ability to converse in a way that feels natural and intuitive*, in which the text would have the same meaning without the first sentences, however, the first sentence provides extra information to the following subject, or *The model has been trained on a wide range of data, including books, articles, and other source of information*, where the first sentence would make sense by itself, but the second sentence further specifies *wide range of data*.

Most of the phrases use active voice; however, there are some instances of passive voice structures. This is illustrated in following examples:

(23) ...*a large language model that **has been trained** on a massive corpus...*  
...*velký jazykový model, který **byl trénován** na rozsáhlém korpusu...*

(24) *This is a concern that **has been raised** by many experts...*  
*Tato obava **je prezentována** mnoha odborníky*

(25) ...*the platform **was designed** to provide valuable information to users....*  
...*platforma **byla navržena** tak, aby poskytovala cenné informace uživatelům...*

As mentioned in previous subchapters, passive voice structures have equivalents in Czech language. However, these examples of passive voice structures are in present perfect and Czech language has only three tenses, past, present, and future. Therefore, there are two options of translations of the examples depending on if it is still happening or if it has ended. Which option will be chosen depends on context. Cases 23 and 24 are examples of substitution because there is no direct equivalent in the target language. The last example is translated as calque because Czech language offers an equivalent.

#### 4.4.3 Text level

From a visual point of view, the article looks textually diverse, however, from an aesthetical point of view, the text needs to be more varied. The article is divided into nine paragraphs, which are four to six rows long, containing 3,225 characters forming 542 words spreading out over two A4 pages. Since the text is generated by the chatbot, there is no information about the author and also no information about the date of publication.

The chatbot itself cannot generate any images, which is the reason why there are not any pictures to make the article more aesthetically pleasing. Moreover, it cannot use different font types or sizes. Given that, the article has the same formatting everywhere, including the title and headings. On the other hand, there is a title and five headings which divide the text into a more acceptable and cohesive form.

Finally, there is quite a few cohesive devices appearing in the text. Firstly, there are nouns and pronouns referring to *ChatGPT*, such as *it*, *the platform*, or *language model*. These

reference words connect ideas and usually refer back to previously mentioned concepts. Secondly, there are conjunctions connecting related ideas and creating smooth transitions between sentences and paragraphs. Lastly, the article contains a relatively high number of repetitions making the text more lexically sparse. Repeated terms are mainly *ChatGPT* and its synonyms or reference words, however, there are also very similar sentences mentioning the fact that ChatGPT is trained on a large corpus of data. The following sentences illustrate that repetition:

1. *ChatGPT is a large language model that has been trained on a massive corpus of text data, The model has been trained on a wide range of text data, including books, articles, and other sources of information.*
2. *Because the platform has been trained on such a large corpus of data, it can quickly process large amounts of information and provide well-informed responses*
3. *Because the platform is trained on a large corpus of data, it may pick up biases that exist in that data.*

The fact that the chatbot used a single sentence over and over again might create a feeling that it really wants to stress out that information, or it is still not able to generate a human-like article and, therefore, unintentionally makes the article look disturbing to people.

#### 4.4.4 Pragmatic level

According to Jakobson's language functions division (Zmrzlá, 2022), the text has a referential function with the intention to be informative. The generated text shows instances of conative function because the chatbot tends to bring up achievements and benefits of ChatGPT more than its drawbacks. Moreover, the whole article is more positively expressive, which is definitely creating some impact on the reader. There are also instances of metalingual functions because the article contains information about language generation. AI also repeats some phrases quite often. The repetition of specific phrases or sentences stresses out the information contained in there. Lastly, the chatbot uses the rule of three, which is shown as follows: *books, articles, and other sources of information, or fast, accurate, and informative.*

The quite interesting aspect of the article is that the chatbot uses instances of modality. The following examples illustrate the use of modality:

(26) *ChatGPT has the ability to converse...*  
*ChatGPT má schopnost konverzovat...*

(27) *...it could also be used to spread misinformation...*  
*...mohlo by to také být využito k šíření dezinformací...*

In example 26, AI writes *ChatGPT has the ability to converse...* instead of *ChatGPT can converse*, which shows that the chatbot wants to stress out that information, and in example 27 the chatbot implies the possibilities of different outcomes.

The translation of modality should keep the intention of stressing out information or author's opinion on possibilities of misuse. The first example is *has the ability to*, which could be substituted by *can*, however, that would not stress out the information. The Czech translation would be *má schopnost*, which could also be substituted by *umí*, however, same as in previous case, it would not stress out the information. The second example is *could also be* which can be translated as *by mohlo také být* which still reflects authors assumption and worries. Both cases are Czech equivalents, therefore, the translation method is calque.

#### 4.4.5 Conclusion

In conclusion, the first part of the analysis shows that the most frequent word classes are nouns, prepositions, verbs, and adjectives. There are also some examples of internationalisms, professional terminology of the language of science and technology, and slang terms within the same register. Due to the highly occurrent repetition of non-terminological terminology, where terms can be substituted by synonyms, the text is rather lexically sparse. Secondly, in the syntax level of the analysis, it is apparent that the chatbot followed the standard structure of the sentences in the language of science and technology. The sentences in the article are primarily complex. Moreover, there are some instances of parenthetical and contrastive sentences. The majority of the text is written in active voice, however, there are few instances of passive voice structures. Thirdly, from a visual point of view, the text looks diverse and sparse simultaneously. The text is diverse in terms of division with headings, where there are five headings in that relatively small text, which creates better readability and coherence of the text. However, since the chatbot cannot generate images and use other types of fonts, the text looks plain. The chatbot also used



cohesive devices in the text. Lastly, pragmatically, the article has a referential function with some instances of conative and metalingual functions; and there are also some instances of modality.

## 4.5 Comparison

### 4.5.1 Lexical level

This part contains examples of professional words in the most frequent word classes. However, some words appeared only once in the text. By not using professional verbs and adjectives repeatedly and choosing non-professional verbs instead, authors aim to address a broader audience. That means, that the articles are not primarily professional but popular scientific.

From a lexicological point of view, all four texts are rather similar. The most used words are *ChatGPT* and its synonyms, *OpenAI*, *chat*, *bot*, *users*, *language*, *AI*, and *model*. Those are words appearing among the most frequent words in all four texts, and all of them are nouns. The only difference is a verb use in the fourth text. The most frequent word class are nouns in all four texts. Following nouns in terms of order are verbs, adjectives, and prepositions.

The most used words							
Text 1	n	Text 2	n	Text 3	n	Text 4	n
GPT-4	15	Chat, ChatGPT	7	ChatGPT	26	ChatGPT	18
Users, OpenAI	9	Bot, Language	6	GPT-4	12	Use	9
Model	8	Users, Tools, AI	4	AI	12	Information	7

Table 9 The most used words in all four texts

The most used word classes			
Text 1	Text 2	Text 3	Text 4
Nouns	Nouns	Nouns	Nouns
Adjectives	Verbs	Verbs	Verbs
Verbs	Adjectives	Adjectives	Adjectives

Table 10 The most used word classes in all four texts

The most frequently used nouns are also usually the same in all four texts. These words are the name of the product, which were the articles about *ChatGPT*, the name of the company which created the product, *OpenAI*, and synonymous words for the word *ChatGPT*, such

as *chat*, *bot*, *chatbot*, *AI*, and *model*. Apart from these words are words *version*, *system*, *language*, and *users*. The different most frequent words are not that frequent in other texts are *tools* in the text 2, and *information* and *responses* in the text 4, which was generated by AI. Human-written texts contained more synonyms for the chatbot whereas AI-generated text repeats the name of the product and is not that lexically diverse. The second and third texts also have word *language* as second most appearing, which could mean authors were more interested in the way the chatbot speaks. The first and second texts have the word *users* among the most used words. This implies that the authors might have wanted to speak more about the benefits for the users of the chatbot. The first and third texts also contain the word *version* often because these articles are about the newest version of ChatGPT.

The most used nouns							
Text 1	n	Text 2	n	Text 3	n	Text 4	n
Users	9	Chat	7	AI	12	Information	7
Model	8	Bot, Language	6	Language	8	Data, Responses	6
Version, System	6	Users, Tools, AI	4	Version	7	Language, Model	5

Table 11 The most used nouns in all four texts

The first and third texts have the adjective *new* as the most occurring, which means that the text is dealing with the newest update or the most current version. In contrast, the second and fourth texts seem to be speaking about ChatGPT in general because the most appearing words are *large* and *wide*. That means the authors bring up the amount of source data more than the authors of other articles. Focusing on the most appearing adjective in terms of the language of science and technology, all texts contain word *artificial* in their most frequent words. That is logical because the topic is AI. The first text has the words *capable* and *creative* among the first three most appearing words, which means that author brings up the positive abilities of the chatbot. Text 2 is rather lexically diverse because adjectives in the register of science and technology are not repeating. The third text uses words *worded* and *powerful* often. Lastly, in the fourth text, AI seems to be bringing up the accuracy and naturality of ChatGPT outputs compared to human answers.

Examples of professional adjectives							
Text 1	n	Text 2	n	Text 3	n	Text 4	n
Capable	2	Viral	1	Worded	4	Accurate	4
Artificial	2	Instant	1	Artificial	2	Natural	3
Creative	2	Quantum	1	Powerful	2	Artificial	2

Table 12 Examples of professional adjectives in all four texts

The most frequent professional verb that is appearing in more texts is *create*, which is expectable given that the chatbot is a tool for creating. The first text repeats word *improve*, which points out improvements in the new version. The other verb that appears more often is *submit*. The second does not contain more appearances of any professional verb. The third text uses verbs *generate* and *demonstrate* more than once. Lastly, the text generated by AI uses *provide* often, which brings up the benefits to the users. Other professional verbs are not repeating.

Examples of professional verbs							
Text 1	n	Text 2	n	Text 3	n	Text 4	n
Submit	2	Experiment	1	Create	5	Provide	7
Create	2	Simulate	1	Generate	2	Generate	1
Improve	2	Evoke	1	Demonstrate	2	Explore	1

Table 13 Examples of professional verbs in all four texts

The articles use language specific to the register of science and technology. All four include internationalisms, vocabulary specific to the register, and words in technical/IT slang. Internationalisms which are the same for all four texts are *input/output*, *data*, *model*, *AI*, or *internet*. Terms specific to the register of science and technology are *AI*, *data*, *input/output*, *platform*, *application*, *software*, *demo*, or *plug-in*. Lastly, among the most frequent slang terms in the articles belong *prompt*, *language model*, *bot*, *app*, or *demo*.

#### 4.5.2 Syntax level

The syntax of the articles is also relatively similar, thus, following standard grammatical structures. Sentences are mostly long and complex, which is typical for the language of science and technology. Authors use complex sentences usually to convey more detailed information about the matter or present multiple ideas in a single sentence. Human-written articles also include citations in their complex sentences, usually to share other opinions or to prove facts. There are also some instances of contrastive sentences in all four texts,

which usually serve to reflect drawbacks after highlighting a benefit. Given that AI is a very controversial current topic, there is no wonder that authors try to provide all perspectives about the matter. The last apparent type of phrase is parenthetical, which authors usually use to provide additional information or clarify unambiguous phrases. Lastly, there are some instances of passive voice structures in all four texts, which usually serve to emphasize the action rather than the subject. Passive voice structures mostly occur with the word ChatGPT or its synonyms.

#### 4.5.3 Text level

The structure of all four texts consists of a title and division into paragraphs. The title is usually not a phrase but one sentence. Three out of four articles have a picture near the title; however, ChatGPT cannot generate pictures; therefore, the fourth article does not contain any pictures or other media. The first three articles also mention the name of the authors and the date of publication, whereas the fourth article is generated by AI; therefore, it does not have any author.

The third article is the longest one; therefore, it has the best structural division. Apart from the title and division in paragraphs, it has several headings and subheadings and a bulleted list. Moreover, it is more visually entertaining, with shorter paragraphs and more pictures. The first article has a picture and a video, and it looks visually less diverse than the third text; however, it still looks eye-catching. The least visually diverse article is the second one. It has only a title and does not have any headings or subheadings. There is not even any division with different font sizes. Lastly, all human-made articles contain hyperlinks to mentioned topics. The fourth article is harder to analyze since it is AI-made, and ChatGPT is still limited in terms of what it can do. As it was mentioned, it does not have the capability to create images or any other media; however, it still creates text with a very good structure with the help of headings. Title, headings, and texts have all the same font because it generates text only in one font. Moreover, there are no hyperlinks to mentioned topics in the text.

#### 4.5.4 Pragmatic level

All four articles have a referential function, and the texts are informative. Since the articles deal with the topic of language models, another type of language function is metalingual. The third language function appearing in all four texts is conative, however, for different reasons. In the articles concerned about the newest version GPT-4, their intention is to summarize updates into an article. They want readers to read their article instead of watching a live event or reading the update notes. Therefore, they might use language more acceptable to a broader audience. The articles present general information about the chatbot and usually convey a message that ChatGPT is a valuable tool, and it manages to achieve important milestones; however, users should be cautious because it still has many drawbacks.

There are also instances of modality, even in the AI-generated article. These instances of modality usually have the intention to make a statement less expressive or show authors' subjectivity in the form of their opinion, worries, or assumptions. Lastly, the texts are unambiguous since the topic is in the language of science and technology. Given that, the appearance of figurative language is rare. The example of the figurative language appearing in all four texts is that authors consider an inanimate chatbot animate.

## Conclusion

The aim of the practical part was to create the theoretical knowledge needed for the practical part of the thesis. The practical part consists of 2 main chapters – Style of science and technology and Translation of professional texts – with each consisting of 5 subchapters.

In the first part, *The style of science and technology*, I introduced the topic and divided style into five substyles - Scientific (Theoretical), Practically professional, Non-fictional, Educative and Essayistic. Then I distinguished the basic features of the style, specified functions of the style, and described which forms the style acquired. Subsequently, I divided the forms of the text horizontally and vertically. Lastly, I summarized the lexicon and syntax specific to the given style.

The second part was *Translation of professional texts*, where I described the process of translation and divided the process into three stages. The most crucial parts are two sets of translation procedures, the first by Vinay and Darbelnet and the second by Malone. Vinay and Darbelnet distinguished seven translation processes as follows: borrowing, calque, literal translation, transposition, modulation, equivalence, and adaptation. Malone supplemented their classification and distinguished nine translation procedures followingly: equation, substitution, divergence, convergence, amplification, reduction, diffusion, condensation, and reordering. Lastly, I specified terminology, lexicology, and terminology translation.

The middle part of the thesis is the methodology, which explains the objectives and research questions of the analytical part, what the materials are, which are analyzed further on in the analysis, what are the aspects of the analysis, and what procedure will be used.

The aim of the analytical part was to analyze four different texts focusing on a single topic, which will be analyzed on four different levels. This aim was fulfilled. The analysis consists of four levels of analysis of four different texts.

The first level of the analysis is the lexical level, which contains the analysis of vocabulary, if there are any internationalisms, if it uses vocabulary specific to the register of science and technology, and if it uses any slang terms. All four texts contain similar terminology falling within the register of science and technology. Moreover, the part of the lexical level is also the analysis of the most occurring words in the most frequent word classes. The most occurring word class is nouns in all four texts, which is typical in the language of science and technology. Among the most used words is ChatGPT or its synonyms. Considering professional terminology, authors use nouns the most often. Apart from nouns, professional verbs and nouns are usually not used repeatedly and authors chose non-professional words instead. By choosing non-professional verbs and adjectives, authors aim to attract a broader audience, therefore, the text is not professional but popular scientific.

The syntax level contains the analysis of the structure of sentences and overall coherence such as if the texts are straightforward and easy to understand or not. The sentences are mostly complex, which means the authors wanted to present more ideas in a single sentence or provide further explanation of mentioned matters. Given that the topic of AI is controversial these days, authors also use contrastive phrases to provide the most common arguments with their counterarguments to keep the text more objective. Lastly, there are some instances of passive voice structures, which are usually used when speaking about ChatGPT, and the purpose of passive voice is usually to stress out the activity or process instead of the subject.

The third level is the text level. It analyzes the visual structure of the texts. All texts are divided into paragraphs and have titles. Some texts are also divided by headings and subheadings. Only the title, headings, subheadings, and hyperlinks have a different font; otherwise, the rest of the text is written in the same font, except the AI-generated text, which uses the same font in all cases.

Lastly, the pragmatic level contains examples of modality, which occurs in all four cases, situational context, author's intentions, and readers' expectations. All four texts have a

referential function where the texts are meant to be informative. Moreover, the articles also have conative and metalingual functions.

Analyses were usually supplemented with examples of the mentioned phenomena from the articles. Some examples were then translated from English to Czech, and the translation process was analyzed. The majority of translations were translated as a calque because there is an equivalent in the Czech language. Professional terminology or internationalisms are also translated as borrowing.

The last part is the comparison, which only proved the expected outcomes from the analysis. However, the comparison also showed that AI-generated text still needs some elements of human-written texts. The text does not contain images, does not use different fonts, and is less lexically diverse than the other texts. The fact that AI generated text still needs some elements of human-written texts, can be an inspiration for future theses, which can compare human-written texts and AI-written texts. The chatbot would be asked to be more lexically diverse. The future versions might also be able to create images and write in different fonts. When it happens, it might also be interesting to compare it to human-written texts. Lastly, the AI-generated text was generated on a free version of ChatGPT, and it might be interesting to make a comparison of the human-written text with the AI-written text generated by paid GPT-4.



## List of references

- Čechová, M. a kol. (2003). *Současná česká stylistika*. 1. vyd. Praha: ISV nakladatelství
- Čechová, M., Krčmová, M & Minářová, E. (2008). *Současná stylistika*. Vyd. 1. Praha: NLN, Nakladatelství Lidové noviny
- Hrdlička, M. (2010). *Kapitoly o češtině jako o cizím jazyku*. Plzeň: Západočeská univerzita v Plzni
- Jegede, O. O. (2020). *Linguistic Features of English for Science and Technology* (Britain Internaional of Linguistics, Arts and Education Journal). The United Kingdom: BIAR Publisher
- Knittlová, D. (1990). *Funkční styly v angličtině a češtině*. Olomouc: Univerzita Palackého
- Knittlová, D. (1995). *Teorie překladu*. Vyd. 1. Olomouc: Vydavatelství Univerzity Palackého,
- Knittlová, D. (2000). *K teorii i praxi překladu*. Olomouc: Univerzita Palackého
- Knittlová, D, Grygová, B. & Zehnalová, J. (2010) *Překlad a překládání*. 1. vyd. Olomouc: Univerzita Palackého v Olomouci, Filozofická fakulta
- Krhutová, M. (2007). *The language of electrical engineering as a special province*. Brno: Akademické Nakladatelství Cerm®, s.r.o.
- Levý, J. (1958). *Úvod do teorie překladu*. Praha: Státní pedagogické nakladatelství
- Machová, S. (1995) *Terminografie*. Praha: H&H.
- Malone, R.J.L. (1988) *The science of linguistics in the art of translation*. New York: State University of New York Press

Man, O. (2010). Antologie teorie odborného překladu: (výběr z prací českých a slovenských autorů). 3., aktualiz. a dopl. vyd., Na OU 2. Ostrava: Ostravská univerzita v Ostravě

Mitter, P. (2019). Lexikologie a slovtvorba češtiny s úkoly pro kombinované studium. Ústí nad Labem: Univerzita J. E. Purkyně v Ústí nad Labem

Poštolková, B. (1983). O české terminologii. Praha: Academia

Translation Bureau of Public Works and Government Services Canada (2001). Handbook of terminology. Hull: CANADA

Vinay, J.P. & Darbelnet, J. (1995). Comparative Stylistics of French and English: A Methodology for Translation, Amsterdam/Philadelphia: John Benjamins

Zmrzlá, P. (2022). Discourse Under Investigation: A Brief Summary of Theoretical Background and Concepts.

Žváček, D. (1995) Kapitoly z teorie překladu I: (odborný překlad). 1. vyd. Olomouc: Vydavatelství Univerzity Palackého

Online sources:

Britannica. Kinds of dictionaries. <https://www.britannica.com/topic/dictionary/Kinds-of-dictionaries>

Sources of the articles:

Hern, A., & Bhuiyan, J. (2023). OpenAI says new model GPT-4 is more creative and less likely to invent facts. *The Guardian*.

<https://www.theguardian.com/technology/2023/mar/14/chat-gpt-4-new-model>

Hughes, A. (2023). ChatGPT: Everything you need to know about OpenAI's GPT-4 tool. *www.sciencefocus.com*. <https://www.sciencefocus.com/future-technology/gpt-3/>

Sundar, S. (2023). If you still aren't sure what ChatGPT is, this is your guide to the viral chatbot that everyone is talking about. *Business Insider*.

<https://www.businessinsider.com/everything-you-need-to-know-about-chat-gpt-2023-1>

## Extended abstract

Cílem bakalářské práce je vypracovat kvalitativní analýzu odborných textů. Práce se skládá z teoretické části, metodiky a praktické části.

Cílem teoretické části je popsat základy odborného stylu, a především stylu vědy a techniky. Zde jsem rozdělil styl na pět typů: vědecký, prakticky odborný, literaturu faktu, naučný a esejistický. Další část popisuje základní rysy, kterými se tento styl vyznačuje, specifikuje, jaké jsou jeho funkce, a popisuje formy, kterých tento styl může nabývat. Následně se forma textu rozdělila na horizontální dělení a vertikální. V neposlední řadě se v této části vyskytuje nástin toho, jak by měla vypadat slovní zásoba a struktura vět.

Druhá část teoretické části se zabývá převážně překladem odborných textů. Na začátku druhé části jsou stručně uvedené nejznámější překladatelské postupy od Vinaye a Darbleneta, které jsou pak doplněné překladatelskými postupy ve verzi od Maloneho. Uvedené typy překladů podle Vinaye a Darbleneta jsou: výpůjčka, kalkování, doslovný překlad, transpozice, modulace, ekvivalence a adaptace. Malone uvedl následující postupy: ekvivalence, substituce, divergence, konvergence, amplifikace, redukce, difúze, kondenzace a přestavba. Dále je popsána terminologie, lexikologie a je vysvětlen rozdíl mezi těmito dvěma obory. Poslední podkapitola v teoretické části se zabývá překladem termínů.

Prostřední částí práce je metodika. V této části jsou vysvětleny cíle analýzy, stanoveny výzkumné otázky, ukázány aspekty analýzy a popsáno, jakým způsobem se bude postupovat. Také je zde seznámení s materiály, které se následně budou analyzovat v praktické části.

Další částí je analytická část jejíž cílem je analyzovat čtyři různé texty, které se zabývají jedním tématem. Každý z těchto textů měl být analyzován na čtyřech úrovních, což bylo splněno.

První úroveň analýzy je lexikální část. Tato část rozebírá slovní zásobu daného textu, jestli se vyskytují v textu nějaké internacionalismy, jestli slovní zásoba odpovídá registru vědy a

techniky nebo jestli autoři používají nějaké slangové výrazy. Všechny čtyři texty obsahovaly podobné termíny, které spadají do registru vědy a techniky. Další částí lexikálního rozboru byla také analýza nejčastěji používaných slov, nejčastěji používaných slovních druhů a nejčastějších slova v jednotlivých slovních druzích. Výsledkem toho bylo zjištění, že všechny čtyři texty nejvíce používaly podstatná jména, která obsahovala i nejvíce použití terminologických výrazů, což je typické pro jazyk vědy a techniky. Mezi nejčastěji používanými slovy byl vždy ChatGPT a jeho synonyma jako chatbot, jazykový model nebo odkaz na jeho nejnovější verzi GPT-4. V rámci sloves a přídavných jmen v registru vědy a techniky nebylo moc slov, které by se často opakovaly, čímž se vytvořila větší lexikální rozmanitost.

Syntaktická úroveň obsahovala analýzu větné stavby a celkové koherence. Příkladem koherence může být, zdali je text díky větné stavbě jednoduchý a srozumitelný. Všechny čtyři texty obsahují převážně komplexní věty, ve kterých se autoři především snažili prezentovat více myšlenek v jedné větě anebo chtěli dovysvětlit pojmy a témata, které ve větě zmínili. Vzhledem k tomu, že téma umělé inteligence je poněkud kontroverzní, autoři používají často kontrastivní fráze, aby poskytli základní argumenty společně s protiargumenty. To by mělo vést k větší objektivitě textu. V neposlední řadě, texty obsahují situace, kde autoři použili pasivní větné struktury. Trpný rod se používal především v souvislosti s chatbotem ChatGPT, kdy autoři chtěli vyzdvihnout aktivitu namísto podmětu.

Třetí úroveň analýzy textů byla textová úroveň. V této úrovni byla rozebírána spíše vizuální stránka textů. Všechny texty byly rozděleny do odstavců a mají název. Některé texty byly ale i rozděleny pomocí nadpisů a podnadpisů. Z hlediska typu písma, všechny odstavce používají stejný font a stejnou velikost písma. Autoři používají jinou velikost písma pouze u nadpisů a podnadpisů. Texty nepsané umělou inteligencí většinou obsahovali i odkazy na články o tématech zmíněných v textu, které byly vyznačeny podtrhnutím nebo jinou barvou písma. Tyto články také používaly obrázky a v jednom textu bylo dokonce i video. Článek vygenerovaný umělou inteligencí byl sice členěn pomocí nadpisů a podnadpisů, každopádně neumí psát jinými velikostmi nebo druhy písma a ani neumí generovat obrázky.

Poslední úroveň byla pragmatická část, ve které se zkoumala modalita, situační kontext, záměry autora a očekávání čtenáře. Všechny čtyři texty obsahovali případy modality, kde se autoři snažili snížit důraz na dané slovo, udělat ho méně expresivní, anebo se pomocí modality snažili vyjádřit svůj názor. Autoři svým názorem většinou poukazovali na pravděpodobnost, za které se daná situace může stát a nebo se snažili vyjádřit vlastní obavy a výhrady vůči danému tématu. Všechny texty mají také referenční funkci a jsou informativní. Kromě zmíněné funkce, se ve všech textech nachází funkce metalingvistická, jelikož se mluví u jazykovém modelu a způsobu, kterým tvoří text, a funkce konativní, jelikož autor se vždy snaží mít nějaký dopad na čtenáře daného článku.

Úplně poslední částí analýzy je porovnání, ve kterém se potvrdily výsledky, které se očekávaly v průběhu analýzy. Všechny texty používají převážně stejnou slovní zásobu, obsahují termíny jazyka vědy a techniky, internacionalismy a slangové výrazy. Texty také používají převážně komplexní větnou stavbu a autoři občas používají pasivní větné struktury. Z vizuálního hlediska jsou texty podobné a rozdělené do paragrafů. Texty také obsahují známky modality. Tyto všechny aspekty splňoval i text vygenerovaný umělou inteligencí, čímž se podobal textům napsaný novináři. Porovnání ale ukázalo, že umělá inteligence stále potřebuje používat určité prvky, které se používají ostatních textech napsaných novináři. Vygenerovaný článek neobsahuje ani jeden obrázek, nepoužívá jiné typy písma a není tak lexikálně rozmanitý, jako jsou ostatní texty. A právě to, že vygenerovanému textu stále chybí určité lidské prvky, může být inspirací pro budoucí práce, které mohou porovnávat texty psané lidmi s texty psané umělou inteligencí. Věci, které byly vytknuty mohou být zahrnuty do požadavku pro umělou inteligenci, aby se tomu předešlo. Umělá inteligence může být tedy požádána, aby vytvořila text, který je lexikálně rozmanitější. Další verze GPT budou třeba moci vytvářet obrázky a používat různé typy písma. Až se tak stane, tak by to také mohl být dobrý námět pro psanou práci a porovnat to s texty, které byly napsány lidmi. Na závěr, text, který jsem si nechal vygenerovat umělou inteligencí, byl vygenerován na bezplatné verzi chatbotu ChatGPT. Mohlo by tedy být zajímavé porovnat texty psané lidmi s texty, které vygeneruje nejnovější placená verze GPT-4.

## Appendices

Article by The Guardian

The Guardian - [OpenAI says new model GPT-4 is more creative and less likely to invent facts | ChatGPT | The Guardian](#)

Alex Hern and Johana Bhyiyan – Mar 14, 2023

### **OpenAI says new model GPT-4 is more creative and less likely to invent facts**

**Latest version can take images as inputs and improves upon many of the criticisms users had, but will still ‘hallucinate’ facts**

The artificial intelligence research lab OpenAI has released GPT-4, the latest version of the groundbreaking AI system that powers ChatGPT, which it says is more creative, less likely to make up facts and less biased than its predecessor.

Calling it “our most capable and aligned model yet”, OpenAI cofounder Sam Altman said the new system is a “multimodal” model, which means it can accept images as well as text as inputs, allowing users to ask questions about pictures. The new version can handle massive text inputs and can remember and act on more than 20,000 words at once, letting it take an entire novella as a prompt.

The new model is available today for users of ChatGPT Plus, the paid-for version of the ChatGPT chatbot, which provided some of the training data for the latest release.

OpenAI has also worked with commercial partners to offer GPT-4-powered services. A new subscription tier of the language learning app Duolingo, Duolingo Max, will now offer English-speaking users AI-powered conversations in French or Spanish, and can use GPT-4 to explain the mistakes language learners have made. At the other end of the spectrum, payment processing company Stripe is using GPT-4 to answer support questions from corporate users and to help flag potential scammers in the company’s support forums.

“Artificial intelligence has always been a huge part of our strategy,” said Duolingo’s principal product manager, Edwin Bodge. “We had been using it for personalizing lessons

and running Duolingo English tests. But there were gaps in a learner’s journey that we wanted to fill: conversation practice, and contextual feedback on mistakes.” The company’s experiments with GPT-4 convinced it that the technology was capable of providing those features, with “95%” of the prototype created within a day.

During a demo of GPT-4 on Tuesday, Open AI president and co-founder Greg Brockman also gave users a sneak peek at the image-recognition capabilities of the newest version of the system, which is not yet publicly available and only being tested by a company called Be My Eyes. The function will allow GPT-4 to analyze and respond to images that are submitted alongside prompts and answer questions or perform tasks based on those images. “GPT-4 is not just a language model, it is also a vision model,” Brockman said, “It can flexibly accept inputs that intersperse images and text arbitrarily, kind of like a document.”

At one point in the demo, GPT-4 was asked to describe why an image of a squirrel with a camera was funny. (Because “we don’t expect them to use a camera or act like a human”.) At another point, Brockman submitted a photo of a hand-drawn and rudimentary sketch of a website to GPT-4 and the system created a working website based on the drawing.

OpenAI claims that GPT-4 fixes or improves upon many of the criticisms that users had with the previous version of its system. As a “large language model”, GPT-4 is trained on vast amounts of data scraped from the internet and attempts to provide responses to sentences and questions that are statistically similar to those that already exist in the real world. But that can mean that it makes up information when it doesn’t know the exact answer – an issue known as “hallucination” – or that it provides upsetting or abusive responses when given the wrong prompts.

By building on conversations users had with ChatGPT, OpenAI says it managed to improve – but not eliminate – those weaknesses in GPT-4, responding sensitively to requests for content such as medical or self-harm advice “29% more often” and wrongly responding to requests for disallowed content 82% less often.



GPT-4 will still “hallucinate” facts, however, and OpenAI warns users: “Great care should be taken when using language model outputs, particularly in high-stakes contexts, with the exact protocol (such as human review, grounding with additional context, or avoiding high-stakes uses altogether) matching the needs of a specific use-case.” But it scores “40% higher” on tests intended to measure hallucination, OpenAI says.

The system is particularly good at not lapsing into cliché: older versions of GPT will merrily insist that the statement “you can’t teach an old dog new tricks” is factually accurate, but the newer GPT-4 will correctly tell a user who asks if you can teach an old dog new tricks that “yes, you can”.

Article by Business Insider

Business Insider - [Everything You Need to Know About ChatGPT \(businessinsider.com\)](https://www.businessinsider.com/everything-you-need-to-know-about-chatgpt)

Sindhu Sundar Mar 1, 2023

## **If you still aren't sure what ChatGPT is, this is your guide to the viral chatbot that everyone is talking about**

Since OpenAI released its blockbuster bot ChatGPT in November, users have casually experimented with the tool, with even Insider reporters trying to simulate news stories or message potential dates.

To older millennials who grew up with IRC chat rooms — a text instant message system — the personal tone of conversations with the bot can evoke the experience of chatting online. But ChatGPT, the latest in technology known as "large language model tools," doesn't speak with sentience and doesn't "think" the way people do.

That means that even though ChatGPT can explain quantum physics or write a poem on command, a full AI takeover isn't exactly imminent, according to experts.

"There's a saying that an infinite number of monkeys will eventually give you Shakespeare," said Matthew Sag, a law professor at Emory University who studies copyright implications for training and using large language models like ChatGPT.

"There's a large number of monkeys here, giving you things that are impressive — but there is intrinsically a difference between the way that humans produce language, and the way that large language models do it," he said.

Chat bots like GPT are powered by large amounts of data and computing techniques to make predictions to string words together in a meaningful way. They not only tap into a vast amount of vocabulary and information, but also understand words in context. This helps them mimic speech patterns while dispatching an encyclopedic knowledge.

Other tech companies like Google and Meta have developed their own large language model tools, which use programs that take in human prompts and devise sophisticated responses. OpenAI, in a revolutionary move, also created a user interface that is letting the general public experiment with it directly.

Some recent efforts to use chat bots for real-world services have proved troubling — with odd results. The mental health company Koko came under fire this month after its founder wrote about how the company used GPT-3 in an experiment to reply to users.

Koko cofounder Rob Morris hastened to clarify on Twitter that users weren't speaking directly to a chat bot, but that AI was used to "help craft" responses.

The founder of the controversial DoNotPay service, which claims its GPT-3-driven chat bot helps users resolve customer service disputes, also said an AI "lawyer" would advise defendants in actual courtroom traffic cases in real time, though he later walked that back over concerns about its risks.

Other researchers seem to be taking more measured approaches with generative AI tools. Daniel Linna Jr., a professor at Northwestern University who works with the non-profit Lawyers' Committee for Better Housing, researches the effectiveness of technology in the law. He told Insider he's helping to experiment with a chat bot called "Rentervention," which is meant to support tenants.

That bot currently uses technology like Google Dialogflow, another large language model tool. Linna said he's experimenting with Chat GPT to help "Rentervention" come up with better responses and draft more detailed letters, while gauging its limitations.

"I think there's so much hype around ChatGPT, and tools like this have potential," said Linna. "But it can't do everything — it's not magic."

OpenAI has acknowledged as much, explaining on its own website that "ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers."

Article by BBC Science Focus

[ChatGPT: Everything you need to know about OpenAI's GPT-4 upgrade | BBC Science Focus Magazine](#)

Alex Hughes - 03rd April, 2023

## **ChatGPT: Everything you need to know about OpenAI's GPT-4 tool**

**OpenAI is back in the headlines with news that it is updating its viral ChatGPT with a new version called GPT-4. But when will this be available, how does it work and can you use it?**

OpenAI has quickly become one of the biggest names in tech. The artificial intelligence (AI) company has made realistic image generators, 3D-model creators and, the thing it is now best-known for, ChatGPT.

With the power to pass legal exams, write entire feature-length articles, and even code full websites, ChatGPT has people talking about the power of AI.

Now, the company has announced a new major upgrade to the software behind ChatGPT. While the program has been running on technology known as GPT-3, OpenAI has now officially launched GPT-4.

While they are not exactly catchy names, GPT-3 and now 4 are actually the internet's best-known language-processing AI models. Since ChatGPT was announced, it has been banned in select schools and utilised by major companies like Microsoft.

Now, the company has also introduced a pay-to-use version called ChatGPT Pro. This offers users a host of added benefits for \$20 (£16) a month, including priority access, faster load times, and now access to GPT-4.

So what is ChatGPT? What does it do? And is this really the future of AI? We've answered these questions and more down below.

## **What is GPT-3, GPT-4 and ChatGPT?**

GPT-3 (Generative Pretrained Transformer 3) and GPT-4 are state-of-the-art language processing AI models developed by OpenAI. They are capable of generating human-like text and have a wide range of applications, including language translation, language modelling, and generating text for applications such as chatbots. GPT-3 is one of the largest and most powerful language processing AI models to date, with 175 billion parameters.

Its most common use so far is creating ChatGPT - a highly capable chatbot. To give you a little taste of its most basic ability, we asked GPT-3's chatbot to write its own description as you can see above. It's a little bit boastful, but completely accurate and arguably very well written.

In less corporate terms, GPT-3 gives a user the ability to give a trained AI a wide range of worded prompts. These can be questions, requests for a piece of writing on a topic of your choosing or a huge number of other worded requests.

Above, it described itself as a language processing AI model. This simply means it is a program able to understand human language as it is spoken and written, allowing to understand the worded information it is fed, and what to spit back out.

## **How much does ChatGPT cost and how can you use it?**

ChatGPT is very easy to sign up for and use, simply:

head over to the ChatGPT website and create an account.

You'll need to wait until your account has been accepted (you can skip this step if you have an account from Dall-E 2).

Logging in will present you with a very simple page. You are offered some example prompts, and some information about how ChatGPT works.

At the bottom of the page is a text box. This is where you can ask ChatGPT any of your questions or prompts.

Currently, ChatGPT remains free-to-use software. However, OpenAI has now announced ChatGPT Pro - a pay-to-use version with added benefits.

This version of the software will cost \$20 (£16) a month, offering users priority access, quicker load times and access to updates and new features before anyone else.

For now the free version remains but it is unclear if this will change in the future.

### **How is GPT-4 different to GPT-3?**

In essence, GPT-4 is the same as its predecessor GPT-3. However, there are some new features that boost the software's abilities.

Mainly, GPT-4 includes the ability to drastically increase the number of words that can be used in an input... up to 25,000, 8 times as many as the original ChatGPT model.

Equally, OpenAI has stated that the latest version of their technology makes fewer mistakes that they are calling 'hallucinations'. Previously, ChatGPT could become confused, offering up a nonsensical answer to your question, or even inputting stereotypes or false information.

Additionally, GPT-4 is better at playing with language and expressing creativity. In OpenAI's demonstration of the new technology, ChatGPT was asked to summarise a blog post only using words that start with the letter 'g'. It also has a better understanding of how to write poetry or creative writing, but it is still by no means perfect.

On top of this, OpenAI also displayed the potential of using images to initialise prompts. For example, the team showed an image of a fridge full of ingredients with the prompt "what can I make with these products?". ChatGPT then returned a step-by-step recipe.

While it wasn't demonstrated, OpenAI is also proposing the use of video for prompts. This would, in theory, allow users to input videos with a worded prompt for the language model to digest.

Creating recipes with images is a clever use of the technology, but it is only the tip of how images could be used with ChatGPT. The company also demonstrated the ability to create a whole website that successfully ran JavaScript with just a handwritten sketch of a website.

As a tool to complete jobs normally done by humans, GPT-3 was mostly competing with writers and journalists. However, GPT-4 is being shown to have the ability to create websites, complete tax returns, make recipes and deal with reams of legal information.

### **Concerns for the future**

A general apprehension has followed artificial intelligence throughout its history and things are no different with ChatGPT. Critics have been quick to raise the alarms over this technology, but now even those closest to it are utilising caution.

An open letter has been drafted calling for all AI labs to pause for at least six months on the development of systems more powerful than GPT-4. This would include OpenAI's work on GPT-5 - the next version of technology ChatGPT will eventually run on.

This open letter has been signed by prominent AI researchers, as well as figures within the tech industry including Elon Musk, Steve Wozniak and Yuval Noah Harari.

This letter states that the pause should be public and verifiable, arguing that companies like OpenAI, Microsoft and Google are entering a profit-driven race to develop and release new AI models at a dangerous pace.

This comes at the same time as a report from Goldman Sachs that suggested 300 million full-time jobs could be impacted by AI systems like ChatGPT, escalating existing concerns around these platforms.

Along with both of these events, Italy also became the first country to ban ChatGPT. This was, however, due to concerns over the processing and storage of data used to train the platform. While Italy is the only country to take such action so far, both the UK and a collection of EU countries have raised concerns around the platform.



## Article generated by AI – Chat GPT

Source Chat GPT – <https://chat.openai.com/chat> - 8 April, 2023

Questions: „Write a news article about Chat GPT. Tell me what it does and its major advantages and drawbacks. The length of the article should be 2 A4 pages.“

„Imagine you issue the article in the news. How would you name it?“

### Chat GPT: The Revolutionary AI Language Model That's Changing Information Access

Artificial intelligence has made some incredible strides over the past few years, and Chat GPT is no exception. As a large language model trained by OpenAI, Chat GPT has the ability to converse with humans in a way that feels natural and intuitive. In this article, we'll explore what Chat GPT is, how it works, and its major advantages and drawbacks.

#### What is Chat GPT?

Chat GPT is a large language model that has been trained on a massive corpus of text data. This allows it to understand and respond to natural language input from humans. For example, if you ask Chat GPT a question like "What is the capital of France?" it will be able to provide you with a quick and accurate response.

#### How does Chat GPT work?

Chat GPT is based on the GPT-3.5 architecture, which is a deep learning model that uses a transformer-based neural network to process and understand natural language input. The model has been trained on a wide range of text data, including books, articles, and other sources of information.

To use Chat GPT, you simply need to provide it with a prompt or question, and it will generate a response based on its understanding of the context and the information available

in its training data. The platform can handle a wide range of topics and can provide detailed, accurate responses to complex questions.

### Advantages of Chat GPT

One of the primary advantages of Chat GPT is its ability to provide fast, accurate, and informative responses to a wide range of questions. Because the platform has been trained on such a large corpus of data, it can quickly process large amounts of information and provide well-informed responses.

Another advantage of Chat GPT is its versatility. It can be used in a wide range of industries, including healthcare, finance, education, and more. For example, it can be used in healthcare to help patients find information about their symptoms and possible treatment options. It can also be used in finance to help investors make informed decisions about their investments.

### Drawbacks of Chat GPT

One of the main concerns surrounding Chat GPT is the potential for bias in its responses. Because the platform is trained on a large corpus of data, it may pick up biases that exist in that data. This could lead to inaccurate or unfair responses that reinforce existing biases.

Another potential drawback of Chat GPT is the potential for misuse. While the platform was designed to provide valuable information to users, it could also be used to spread misinformation or to manipulate public opinion. This is a concern that has been raised by many experts, and it highlights the need for responsible use of the technology.

### Conclusion

Chat GPT is an impressive achievement in the field of artificial intelligence. Its ability to provide accurate and informative responses to a wide range of questions makes it a valuable tool in many industries. However, as with any technology, it is important to consider the potential drawbacks and to use it responsibly. With careful consideration and

responsible use, Chat GPT has the potential to be a game-changer in the world of information access and communication.