COMENIUS UNIVERSITY IN BRATISLAVA

FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

EXPLORING THE ROLE OF CHATGPT IN CZECH CULTURE

Diploma thesis

2023

Klára Petrovická, BSc

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

EXPLORING THE ROLE OF CHATGPT IN CZECH CULTURE

Diploma thesis

Study program:	Cognitive Science
Field of study:	Computer Science
Department:	Department of Applied Informatics
Supervisor:	doc. RNDr. Martin Takáč, PhD.

Bratislava 2023 Klára Petrovická, BSc

48153355

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Klára Petrovická
Study programme:	Cognitive Science (Single degree study, master II. deg., full time form)
Field of Study:	Computer Science
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak
Title:	Exploring the role of ChatGPT in Czech culture
Annotation:	Since the large language model ChatGPT was introduced in November 2022, it immediately became a viral sensation across various fields and communities. Recent studies demonstrated that ChatGPT can achieve great achievements in a broad range of NLP tasks, however, not so much has been debated in regards its societal implications. By setting the ChatGPT into the Czech cultural setting, this thesis aims to explore and discuss the applied potentials and concerns of this recently introduced cultural sensation.
Aim:	1. Conduct an interdisciplinary review of the ChatGPT large language model from various perspectives, e.g., machine learning, psychology (attention), neuroscience. 2. Theoretically analyse the shortcomings of large language models and the potentials of ChatGPT for the field. 3. Explore how the ChatGPT will shape different aspects of culture and society, e.g., healthcare, education, job market (applied to Czech context).
Literature:	1. Shen, Y., Heacock, L., Elias, J., Hentel, K. D., reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. Radiology. Advance online publication. https://doi.org/10.1148/radiol.230163 2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.
Supervisor:	doc. RNDr. Martin Takáč, PhD.
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	doc. RNDr. Tatiana Jajcayová, PhD.
Assigned:	01.03.2023
Approved:	01.03.2023

prof. Ing. Igor Farkaš, Dr.
Guarantor of Study Programme

……………………………… ………………………
Student Supervisor

48153355

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Klára Petrovická
Študijný program: kognitívna veda (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Exploring the role of ChatGPT in Czech culture
Skúmanie miesta ChatGPT v českom prostredí

Anotácia: Veľký jazykový model ChatGPT sa od svojho spustenia v novembri 2022 okamžite stal virálnou senzáciou v rôznych oblastiach a komunitách. Nedávne štúdie preukázali, že ChatGPT môže dosiahnuť veľké úspechy v širokom spektre úloh NLP, avšak o jeho spoločenských dôsledkoch sa toľko nediskutovalo. Zasadením ChatGPT do českého kultúrneho prostredia si táto práca kladie za cieľ preskúmať a prediskutovať aplikačné možnosti a problémy tejto nedávno predstavenej kultúrnej senzácie.

Cieľ: 1. Prezentovať interdisciplinárny prehľad veľkého jazykového modeluChatGPT z rôznych hľadísk, napr. strojového učenia, psychológie (pozornosti), neurovedy.
2. Teoreticky analyzovať nedostatky veľkých jazykových modelov a potenciálChatGPT pre túto oblasť.
3. Preskúmať, ako bude ChatGPT formovať rôzne aspekty kultúry a spoločnosti,napr. zdravotníctvo, vzdelávanie, trh práce (aplikované na český kontext).

Literatúra: 1. Shen, Y., Heacock, L., Elias, J., Hentel, K. D., reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. Radiology. Advance online publication. https://doi.org/10.1148/radiol.230163
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.

Vedúci: doc. RNDr. Martin Takáč, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. RNDr. Tatiana Jajcayová, PhD.

Dátum zadania: 01.03.2023

Dátum schválenia: 01.03.2023

prof. Ing. Igor Farkaš, Dr.
garant študijného programu

....
študent

....
vedúci práce

Acknowledgment

I realize how fortunate I am to conduct my thesis on a topic I really care about. Countless interactions—minor and significant—with brilliant individuals all shaped the thought lines presented in this thesis.

Naturally, there is one person to whom I want to thank the most, doc. RNDr. Martin Takáč, PhD., my dear supervisor, who agreed to join this adventurous journey on very short notice. Yet, he expressed even more excitement than I did, for which I am unconditionally thankful. His calm perspective, insightful ideas, feedback, and encouraging words made this thesis possible. I am grateful for all the academic experiences during my studies that I got to be supervised by him.

The close second in my letter of appreciation is my partner in crime, Kuba. I am rarely speechless, but I do not know what to thank him for first. He is the author of the present thesis idea. Without him, this thesis will literally be nonexistent. I also appreciate the endless study support calls, re-re-reading my drafts, and letting me run my ideas by him when I could not move any further in my argumentation. I hope to repay him with the same amount of patience and support next year when it is his turn to write his soul out.

I also must thank my family, who have kept me safe and balanced yet through another deadline. They lived through all my previous deadlines with me, and I can proudly say this was our best one so far. And especially for the one and only: Děkuju, mami.

I must also appreciate the bravery of the four professionals, Katka, Josef, Alžběta, and Eva, who offered their expertise and opinions for the sake of this thesis. I must thank them for the interactions, which were incredibly valuable, insightful but also lovely. I enjoyed the conversations with them immensely.

Many more people should be credited here: Miška, for being a fantastic friend, Slovak household, reviewer, and translator. Thomas and Aleš, for kindly offering me their perspectives on the topic in discussion. Barbora and Brigitte, who both survived my indecisiveness, and even though they lobbied for opposing outcomes, helped me realize what I wanted to do. To other friends and family (and Kuba's family) – thank you for tolerating me, I shall become more bearable again soon.

Finally, I want to thank Visegrad Fund for supporting me in my university studies. I am a proud scholar, and I hope this work makes you proud too.

Abstract

Large language models have recently garnered widespread popularity and significant media attention due to their impressive performance on various tasks. ChatGPT, a chatbot developed by OpenAI, is one such implementation of a large, pre-trained language model that became a viral sensation. Early recent studies demonstrated that ChatGPT could exhibit outstanding performance in a broad range of tasks, such as text generation, question answering, and translation. However, little hard evidence is available regarding its impacts on society. Understanding such societal consequences is essential because it can provide insights into the potential success or failure of ChatGPT and mitigate the risks regarding its safe deployment. Therefore, in this thesis, we conduct an interdisciplinary mixed-method study exploring the role of ChatGPT in specific cultural settings of the Czech Republic. We first theoretically and systematically review the model and its benefits, threats, and artificial general intelligence potential. Then, we perform in-depth qualitative analyses of four experts' interviews and 201 Czech news articles collected over three months to define the main topics about ChatGPT among Czech populations. Our results show that people in the Czech Republic have a low adoption rate of ChatGPT due to a missing regulatory body and fear of novelty. The fear is artificially created by media dominance of topics, such as technological competition, job dismissal, and an apocalypse caused by ChatGPT. One topic exception is education, where overwhelmingly positive sentiment is expressed. We present these findings by providing specific examples for each topic and then discuss the implications of (non)addressing these concerns.

Keywords: ChatGPT, Generative Pretrained Transformer, Czech Republic, Public Awareness, Artificial General Intelligence, AI Ethics

Abstrakt

Veľké jazykové modely si v poslednom čase získali veľkú popularitu a značnú pozornosť médií vďaka ich pôsobivému výkonu v rôznych úlohách. ChatGPT je chatbot vyvinutý spoločnosťou OpenAI a je práve jednou z takýchto implementácií veľkého, vopred natrénovaného jazykového modelu, ktorý sa stal virálnou senzáciou. Nedávne prvé štúdie ukázali, že ChatGPT môže vykazovať vynikajúci výkon v širokom spektre úloh, ako je napríklad generovanie textu, odpovedanie na otázky a preklad medzi jazykmi. O jeho vplyve na spoločnosť je však k dispozícii len málo poznatkov. Pochopenie spoločenských dôsledkov je dôležité, pretože môže poskytnúť informácie o spôsobe úspešného a bezpečného adoptovania ChatGPT spoločnosťou. V tomto článku preto uskutočňujeme interdisciplinárnu štúdiu s dizajnom zmiešaných metód, ktorá skúma úlohu ChatGPT v špecifickom kultúrnom prostredí Českej republiky. Najprv teoreticky a systematicky skúmame model a jeho výhody, hrozby, a jeho potenciál pre umelú všeobecnú inteligenciu. Potom podrobne a kvalitatívne analyzujeme štyri rozhovory s odborníkmi ako aj 201 českých novinových článkov zozbieraných počas troch mesiacov s cieľom definovať hlavné témy spojené s ChatGPT naprieč českou populáciou. Naše výsledky ukazujú, že ľudia v Českej republike majú nízku mieru prijatia ChatGPT pre strach z neznámeho a nedostatok regulačných orgánov. Tieto obavy sú umelo vyvolané mediálnou dominanciou tém zaoberajúcich sa technologickou konkurenciou, prepúšťaním zo zamestnania a apokalypsou, čo sú dôsledky pripisované ChatGPT. Výnimkou je téma vzdelávania, kde prevládajú pozitívne nálady. Tieto zistenia prezentujeme na konkrétnych príkladoch jednotlivých tém a následne diskutujeme o dôsledkoch (ne)riešenia týchto obáv.

Kľúčové slová: ChatGPT, generatívny predtrénovaný transformátor, Česká republika, informovanosť verejnosti, umelá všeobecná inteligencia, etika umelej inteligencie

Contents

1. Introduction 12

1.1. Aims 14

1.2. Research Questions 15

1.3. Methodology 16

2. Transformation of Language Models 17

2.1. History of NLP 19

2.2. Model Architecture 25

2.2.1. GPT-1: Transformers, Self-Attention, and Pre-training 25
2.2.2. GPT-2: Staged Release 28
2.2.3. GPT-3: Reinforcement Learning from Human Feedback and InstructGPT 30
2.2.4. GPT-4: Great Prognosis 32

3. Technological Impact 35

3.1. Potentials 37

3.1.1. Education 39
3.1.2. Healthcare 40
3.1.3. Job Market 41

3.2. Risk Areas 45

3.2.1. Discrimination, Exclusion and Toxicity 45
3.2.2. Misinformation Harms 47
3.2.3. Information Hazards 49
3.2.4. Malicious Uses 50
3.2.5. Human-Computer Interaction Harms 52
3.2.6. Automation, Access, and Environmental Harms 53

3.3. *Super*Intelligence Potential 57

3.3.1. Consciousness 59
3.3.2. Intelligence 61

3.4. Theoretical Conclusion 65

4. Adaptation by Society 66

4.1. Czech Narrative 67

4.2. Experts Interviews – Identifying the Topics of Debates 74

4.2.1. Low Adoption & Implementation Rates 75

4.3. Media Analysis 81

4.3.1. Content Analysis Results 82

4.4. Discussion 89

5. The Final Chapter: Conclusion 92

5.1. Limitations 93

References 95

Appendix A: Malicious Use Example 120

Appendix B: Public Attitudes Towards Technology 121

Appendix C: Techno-attitudes Framework 122

Appendix D: Semi-structured Interview Guide 123

Appendix E: Consent Form 124

Appendix F: List of Supplementary Materials 125

Appendix G: Czech News Articles' Links 126

List of Tables

Table 1 Most Popular Schools of Thought 20

Table 2 Versions of GPT-2 Model 29

Table 3 Origin of the Data Used for Training the GPT-3 Model 31

Table 4 Comparison of Technical Parameters of GPT Series Models 34

Table 5 Themes Identified via Thematic Analysis 75

Table 6 Codes of the News Articles for Content Analysis 82

List of Figures

Figure 1. Generated answer from a web interface ChatGPT-3.5 13

Figure 2. Parameter counts of language models 23

Figure 3. Transformer model architecture 27

Figure 4. ChatGPT awareness among adults 36

Figure 5. Demonstration of four prompting methods 44

Figure 6. Percentage of optimistic public attitudes 68

Figure 7. A framework of a spectrum of attitudes towards technology 69

Figure 8. A proposed framework of artificial intelligence attitudes 72

Figure 9. Illustrative images from four selected Czech news articles 85

Figure 10. Illustration of malicious prompts 120

Figure 11. Percentage of (a) pessimistic and (b) ambiguous public attitudes 121

Figure 12. Techno-attitudes towards artificial general intelligence 122

List of Abbreviations

AI – Artificial Intelligence

AGI – Artificial General Intelligence

API – Application Programming Interface

BCC – Behavioral Correlates of Consciousness

DHI – Digital Humanism Initiative

DIA – Digital and Information Agency

DL – Deep Learning

DNN – Deep Neural Network

EC – European Commission

EU – European Union

FLI – Future of Life Institute

GenAI – Generative Artificial Intelligence

GPT – Generative Pre-trained Transformer

IIT – Integrated Information Theory

IQ – Intelligence Quotient

LLM – Large Language Model

MT – Machine Translation

NCC – Neural Correlates of Consciousness

NLP – Natural Language Processing

OECD – Organization for Economic Co-operation and Development

QA – Question-Answering

RLHF – Reinforcement Learning from Human Feedback

RNN – Recurrent Neural Network

TA – Thematic Analysis

1. Introduction

“Recall, the argument is that we are witnessing not a marriage but a divorce between successful engineered agency and required biological intelligence.” – Floridi and Chiriatti (2020)

The present thesis explores the role of a newly launched ChatGPT in society. ChatGPT[1] is a conversational application programming interface (API) which has caused an enormous response and interest among the general population since its launch in November 2023. ChatGPT, developed by OpenAI, is based on a powerful language model explicitly trained to converse with (human) users in a human-like manner – it is a *chatbot*. A chatbot is a conversational agent that can process input and accordingly provide output through written interaction, imitating natural language dialogue (Huang et al., 2021; Janssen et al., 2021). Based on the user’s instructions, ChatGPT can perform a variety of text-based tasks in a manner unrecognizable from human-written text, such as writing poems, formulating university essays and homework, and exhibiting a sufficient level of code programming (Bubeck et al., 2023; OpenAI, 2023a; Sadasivan et al., 2023). It can also demonstrate its capabilities and introduce itself, as shown in **Figure 1**.

Due to ChatGPT’s language capabilities, people all around the world have rapidly adopted this artificial intelligence (AI) tool. Since it was initially available in 161 countries (OpenAI, n.d.), anyone could reach it via its free and user-friendly API from anywhere apart from China, Russia, Belarus, Venezuela, Afghanistan, Iran, and Ukraine[2]. Therefore, it gained over one million subscribers in only five days (Altman, 2022),[3] leaving behind other famous online services like Netflix[4] and Instagram[5] (Similarweb, 2023). To put the disruptive ChatGPT response in context, for Instagram, it took approximately 2.5 months (70 days longer than ChatGPT) to achieve the milestone of one million users (Buchholz, 2023). Since then, the ChatGPT platform has reached one hundred million active users and an estimated one billion total visits in February 2023 (Similarweb, 2023). These numbers demonstrate that since its introduction, ChatGPT has evoked a worldwide interest, making the hype about AI systems finally *mainstream* (Vincent, 2022).

---

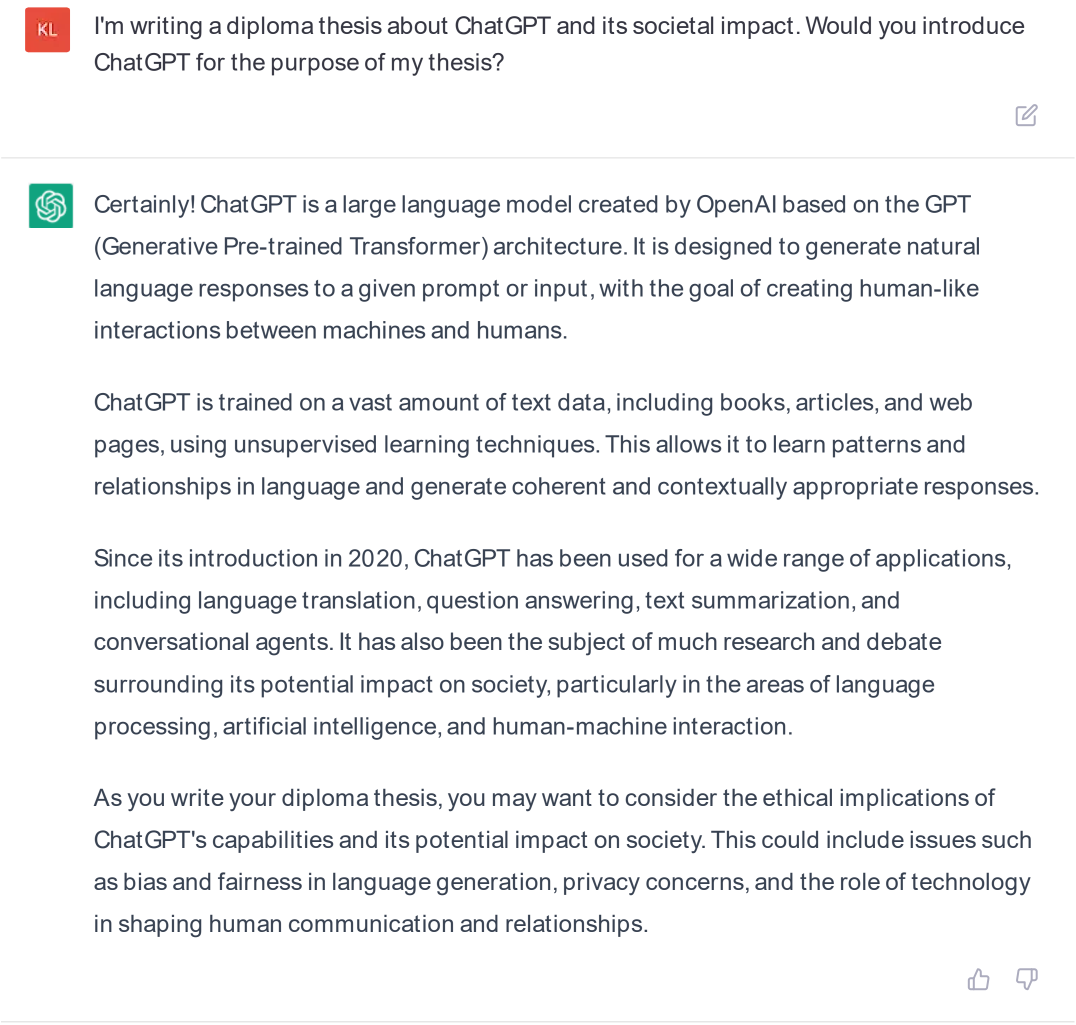
[1] https://chat.openai.com/

[2] Notably, this list is not exhaustive, and more countries cannot access ChatGPT through omission, blocks, or bans. OpenAI, the founding company, maintains a list of countries which may currently use their API: https://platform.openai.com/docs/supported-countries

[3] Information coming from the service itself. Chronologically, this was about five months ago, at the time of the birth of this thesis.

[4] https://www.netflix.com/

[5] https://www.instagram.com/



KL I'm writing a diploma thesis about ChatGPT and its societal impact. Would you introduce ChatGPT for the purpose of my thesis?

Certainly! ChatGPT is a large language model created by OpenAI based on the GPT (Generative Pre-trained Transformer) architecture. It is designed to generate natural language responses to a given prompt or input, with the goal of creating human-like interactions between machines and humans.

ChatGPT is trained on a vast amount of text data, including books, articles, and web pages, using unsupervised learning techniques. This allows it to learn patterns and relationships in language and generate coherent and contextually appropriate responses.

Since its introduction in 2020, ChatGPT has been used for a wide range of applications, including language translation, question answering, text summarization, and conversational agents. It has also been the subject of much research and debate surrounding its potential impact on society, particularly in the areas of language processing, artificial intelligence, and human-machine interaction.

As you write your diploma thesis, you may want to consider the ethical implications of ChatGPT's capabilities and its potential impact on society. This could include issues such as bias and fairness in language generation, privacy concerns, and the role of technology in shaping human communication and relationships.

Figure 1. Generated answer from a web interface ChatGPT-3.5 at https://chat.openai.com/ to a human-given prompt[6] specified above with initials 'KL' on March 14, 2023.

As stated in the generated introduction piece (**Figure 1**), ChatGPT has been the subject of many debates surrounding its potential impact on society. The capabilities of ChatGPT and the hype it caused made a strong impression on researchers worldwide since AI systems are no longer solely an exclusive interest for academics and product developers (Greshake et al., 2023). Consequently, the question that scientists started to ask is how different stakeholders perceive the use of ChatGPT and what is the public discourse about the application (Haque et al., 2022; Leiter et al., 2023; Taecharungroj, 2023; Tlili et al., 2023). Such studies are necessary to identify and define two concepts: expected societal impact and technology adoption rate among the general population (Selwyn & Cordoba, 2022). Although interest was raised, only a few studies looking into the impact of ChatGPT exist. Since ChatGPT has only been around for a couple of months, only one meta-

---

[6] A prompt is a natural language description/instruction of a task.

analysis exploring ChatGPT's current perceptions after 2.5 months (Leiter et al., 2023) and one study conducting an in-depth qualitative sentiment analysis (Haque et al., 2022) are available. Other already existing studies are (a) non-per-reviewed pre-prints, (b) using only Twitter and automatic analysis (e.g., latent Dirichlet allocation) as a data source, (c) lacking an interdisciplinary approach, which is arguably necessary to explore such a complex fundamental topic, and (d) focusing solely on English-speaking users (Haque et al., 2022; Leiter et al., 2023; Taecharungroj, 2023; Tlili et al., 2023).

However, to study such a complex system as ChatGPT, it is necessary to consider the model context with its evolution, its societal impact, and the social environment in which the system operates (Milner, 1981; Rahwan et al., 2019). Therefore, the present study will interdisciplinary and systematically review the ChatGPT's shortcomings and potentials and contribute to the debate regarding its societal implications. This could benefit future regulatory and implementation attempts of a deeper understanding of ChatGPT. Additionally, it aims to identify and define the main topics of debates and concerns regarding ChatGPT among the Czech population. It argues that since ChatGPT is expected to transform society majorly (Elondou et al., 2023), it is crucial to assess the perceptions and concerns of laypeople[7] in a concrete community and context. Due to the historical and cultural specifications of the Czech Republic, such as its central European location, industrial specialization, communist past, and membership in the European Union (EU), the perceptions and implementation rates of ChatGPT are hypothesized to be specific. Thus, this thesis is expected to find a cultural-specific narrative of ChatGPT different from the dominant one present in global (social) media. This should provide a fresh perspective into the debates in a concrete social environment, identifying potential barriers to the safe implementation of the API into the daily lives of the Czech people. This work is the first to comprehensively analyze ChatGPT to this extent and in specific cultural settings, including the Czech perspective.

1.1. Aims

For the above-mentioned reasons, this thesis seeks to achieve three aims through a theoretical and interdisciplinary approach.

1. The first aim is to review the ChatGPT and corresponding LLM from various perspectives, including its model architecture, evolution, and theoretical analysis of its artificial general intelligence potential.

---

[7] The term 'laypeople' will be solely used when referring to *a person or a group of persons who do not have expert knowledge on a particular subject* (AI generated text in particular). For the purpose of this thesis, the terms 'laypeople' and 'general population' are used interchangeably to mean 'non-experts'.

2. Secondly, the present thesis pursues to systematically analyze the threats and potentials of ChatGPT to assess its societal impact.
3. Finally, the present thesis explores, identifies, and defines the main topics of debates and concerns regarding ChatGPT among the Czech population and describes how ChatGPT may shape different cultural aspects in the Czech context.

The overall purpose is to learn more about public discourse regarding the use of ChatGPT from the perspectives of experts[8] and laypeople to understand its broader societal impact. This work should contribute to a more comprehensive public discussion about risks and corresponding strategies for mitigating the identified threats.

1.2. Research Questions

Followingly, to fulfil the above-mentioned aims and research directions, it is desirable to investigate the current situation, the impact, and the future prognosis of ChatGPT (within Czech society). The present thesis addresses these needs, which lead to the following research questions:

RQ1: What is ChatGPT, its artificial general intelligence potential, and societal implications, theoretically reviewed with an interdisciplinary approach?

RQ2: What are the benefits and limitations of implementing ChatGPT into day-to-day lives?

RQ3: What are the perceptions and topics of debates about ChatGPT among the Czech population?

Additionally, given the theoretical premises, it is hypothesized that:

H1: The Czech public discourse about ChatGPT will be close-minded, cautious, and techno-sceptic.

H2: The perception and approach towards ChatGPT among Czech Generative AI-related experts will be regulatory, calling for assessing actions before implementing actions into business or governmental systems (education, healthcare, etc.).

---

[8] Experts hereafter to be seen as *a group of professionals who are very knowledgeable about or skilled on a subject of GenAI*. To be acknowledged, such an expert identification is a rather complicated and partially subjective challenge. Next to the person's background and education, a person's perceived expertise is also based on their media presence or close relations.

1.3. Methodology

To answer the aforementioned aims and research questions, this study adopts a theoretical and qualitative case study approach (Yin, 2009). The theoretical approach applies to aims 1&2, reviewing the ChatGPT from various perspectives and systematically analyzing its concerns and benefits. Exploring the potential downstream social impact that ChatGPT may have, is a challenging aim that demands a deep understanding of both – the technological (eco)system and the social environment (Rahwan et al., 2019). It is hardly possible to fully assess the potential and harms of a system without realizing how it will be deployed within a cultural and historical context. Consequently, this thesis starts with a theoretical exploration of the transformation of human language and language models in **Chapter 2**. Then, **Chapter 3** systematically analyzes the potentials, threats, and concerns related to ChatGPT. The chapter concludes with a discussion about ChatGPT's artificial general intelligence (AGI) potential, touching on the Theory of Mind and arguments from the fields of psychology and neuroscience.

The qualitative part of the thesis in **Chapter 4** benefits from a case study research design (Stake, 1995), which is recommended for exploring phenomena within a specific context (ChatGPT in the Czech Republic). Two data sources were included in the qualitative case study: a thematic analysis of interviews with Czech experts (**Chapter 4.2.**) and a content analysis of the Czech news articles (**Chapter 4.3.**). The chapter is opened with an introduction to general techno-attitudes towards ChatGPT and the Czech social environment, and it is concluded with a discussion of the analyses' findings within the Czech context. The final **Chapter 5** provides closing thoughts on the topic and discusses the thesis limitations.

2. Transformation of Language Models

"Hate having to update my lectures every time a new model comes out, with new examples, making same damn points about hallucination, bias, lack of physical & psychological reasoning etc. Would be great if someone would propose a new architecture that didn't have exactly same flaws." – Gary Marcus

Language is the core of most human communication and interaction. It plays a central role in human thought, forms social and emotional relationships, develops personal and social identity, records knowledge, and communicates social intelligence. It is not merely a tool for achieving shared goals among humans; it is much more than just that (Bommasani et al., 2021). Almost every human society develops spoken and signed languages, which are exceedingly diverse in expression and information structure. Despite this diversity, the world's languages exhibit remarkable agreement in conveying rich meaning (Comrie, 1989).

Thus, language is power. "Language is the place where actual and possible forms of social organization and their likely social and political consequences are defined and contested. Yet it is also the place where our sense of ourselves, our subjectivity, is constructed" (Weedon, 1997, p. 21). Language is, in fact, so powerful that it determines how well a child will do in school and later in life – dominantly based on the number of words the child hears by the age of 3, according to Hart and Risley's (1995) influential longitudinal study. Some parents talk more and use abundant vocabulary and a broad range of gestures in interactions with their children, and such variabilities in language input account for disparities in infants' future development (Hoff, 2003; Huttenlocher et al., 2010; Pan et al., 2005; Rowe & Goldin-Meadow, 2009). The disparity is staggering: where working-class children may hear around 1,200 words per hour from their parents, children from the highest-income families hear approximately 2,100 words per hour. Thus, by age 3, a child from a poor socio-economic home environment would have heard 30 million fewer words[9] than a wealthier child. Consequently, children whose families were on welfare would have had lower measures of language skills at age 9-10 and lower intelligence quotient (IQ) – a score from a set of standardized tests to assess human intelligence (Hart & Risley, 1995).[10]

---

[9] *Words* in the sense of *tokens* – reoccurring concrete instances of a class of objects (such a unique class would be a *type*). In other words, it means the child has heard 30 million concrete instances of repeated words but not 30 million unique words by age 3.

[10] To be transparent, this study raised some controversy due to its alleged built-in racial bias. More can be found at: https://www.npr.org/sections/ed/2018/06/01/615188051/lets-stop-talking-about-the-30-million-word-gap

Language allows us to communicate our intelligence via talking, reading, and writing to others. There is a complicated and complex relationship between language and intelligence. As Czech scientist Michal Pěchouček put it: “Everyone, including us from the research community, shares the tendency to view language as a gateway to intelligence” (Tmejová, 2023)[11]. But also, as the psychologist Steven Pinker viewed it: “Language is the jewel in the crown of cognition” (Pinker, 1994). Additionally, since the ultimate methodological goal of *artificial* intelligence is to acquire the appearance of intelligence[12] (Goldstein & Papert, 1977), it is no surprise that the most significant challenge in AI research has been to comprehend language the way any human would do.

Such a challenge required a specialized field of interest – Natural Language Processing (NLP). NLP aims to computationally comprehend, interpret, and generate human-like language and enable novel and otherwise impossible modes of interaction. NLP has two essential aspects which allow users to communicate in natural language. These aspects of person-machine communication are *understanding* and *generating* language. While natural understanding has been widely studied, the natural language generation has just relatively recently started to be an increasingly active field of research (Paris et al., 2013). In recent years, the improvement and development of large language models (LLMs) have made a remarkable benchmark in AI research thanks to significant progress in language generation (Bubeck et al., 2023; Wu et al., 2023). LLMs digest enormous quantities of text data and infer the relationships between words in a text sequence (Min et al., 2021). Incredible breakthroughs have been made with models such as T5 (Raffel et al., 2019), BLOOM (Scao et al., 2022), and GPT-3 (Brown et al., 2020). However, only GPT has created a cultural disruption.

Although these LLMs significantly contributed to advancements in developing NLP systems that successfully mimic human behavior, there are still notable differences in the linguistic system and the learning process between artificial models and humans. Acknowledging the implications of this gap is crucial to foster a research community to comprehensively understand the potential and limitations of LLMs. Since the evolutionary history of an AI system and its design choices are critical to explaining its behavior (Rahwan et al., 2019), the following sections cover the recent successes of LLMs in NLP and details of how LLMs have changed the overall process of the machine learning modelling of language. Hence, the following **Chapter 2.1.** briefly outlines the history of NLP with important developmental milestones. Then, the architectural design choices of the GPT language

---

[11] Cited from an interview in Forbes.cz (Tmejová, 2023), originally in Czech: “*My všichni, včetně nás z vědecké komunity, máme tendenci vidět jazyk jako bránu k inteligenci.”*

[12] In an overly simplified introductory sense. More regarding the topic of *intelligence* will be discussed later in the thesis in **Chapter 3.3.**

model are described in **Chapter 2.2.** This section also covers the model's training dataset because the nature of the data also influences the system's behavior (Rahwan et al., 2019).

2.1. History of NLP

The ability to work with and process language computationally has been a research interest for an extended period of time. Indeed, in 1950 Alan Turing introduced arguably the most popular test in the history of AI, the Turing Test (Turing, 1950), which targeted a *language*-based interaction between a human and an AI agent[13]. The quest to fulfil the potential of computationally processed language has set the baseline for the new field of NLP, a subfield of linguistics, computer science, and AI. NLP studies the interactions between humans and computers via technologies for the "meaning expressions of words, phrases, sentences, and documents, and for syntactic and semantic processing such as word breaking, syntactic parsers, and semantic parsing and develops applications such as machine translation (MT), question-answering (QA), information retrieval, dialogue, text generation, and recommendation systems" (Zhou et al., 2020, p. 275). Typical challenges in NLP include speech recognition, natural language understanding, and natural language generation.

Nowadays, NLP is crucial for many real-world applications in several aspects of day-to-day life. Conversationally, this increased societal use of NLP technologies is persistently linked to a lack of its general public awareness and understanding (Fast & Horovitz, 2017; Selwyn & Cordoba, 2022). These applications include search engines, customer support systems, business intelligence systems, and virtual assistants. However, over the years, researchers have taken the field in various direction and explored several methodological innovation types. Relatively, it was a long journey for NLP researchers to achieve the current standard, which sets the benchmark of today's models' quality of natural language imitation. It may not have been a long journey timeline-wise because NLP was only introduced approximately 70 years ago and already achieved significant breakthroughs. However, it was a long journey of constantly trying, changing approaches, and reaching dead ends (Cambria & White, 2014). The most important approaches to knowledge representation and NLP research are listed in **Table 1**.[14]

---

[13] AI agent is an autonomously operating system that is designed to perceive its environment, make decisions, and take actions to achieve a specific goal(s).

[14] In the following paragraphs, some of the relevant approaches are described in more detail. However, see Cambria and White (2014) for a more comprehensive outline.

Table 1

Most Popular Schools of Thought in Knowledge Representation and NLP Research

Approach	Characteristics	Reference
Production rule	Cycles of ‘recognize’, ‘resolve conflict’, ‘act’ steps	Chomsky (1957)
Semantic patterns matching	Semantic categories and semantic case frames	Ceccato, 1967
First order logic (FOL)	Axioms and rules of inferences	Barwise (1977)
Bayesian networks	Variables represented by a probabilistic directed acyclic graph	Pearl (1985)
Semantic networks	Patterns of interconnected nodes and arcs	Sowa (1987)
Ontology Web language (OWL)	Hierarchical classes and relationships between them	Mcguinness & Van Harmelen (2004)

Note. Adapted from “ Jumping NLP Curves: A Review of Natural Language Processing Research,” by E. Cambria and B. White, 2014, *IEE Computational Intelligence Magazine*, *9*(2), p. 48 (10.1109/MCI.2014.2307227).

The early focus of NLP in the 1950s was a description of natural language in the form of rule-based methods, also called *production rules* (Chomsky, 1957; Lehnert, 1977; Schank et al., 1973). Such a production rule typically consists of an antecedent set of conditions and a consequent set of actions (Chomsky, 1957). Condition-action pair in the simplest version goes as: IF <condition> THEN <action> (Cambria & White, 2014). These production rule models were already used to build NLP systems such as word/sentence analysis, QA, or MT (Katz et al., 2023). However, there are apparent issues with scalability when the rule-based systems become larger.

In the 1990s, the NLP field (like any other AI field) experienced a severe shift in development. Related to the rapid growth of the internet, large amounts of data became easily available and cheaper due to the decline in the cost of data storage. Additionally, the advent of Moore’s Law (Moore, 1965) and the growing advancements of algorithmic methods have also contributed to NLP becoming a statistical data-driven field (Katz et al., 2023; Shalf, 2020). With human-designed features, statistical models learnt by using labeled data. Due to the statistical learning methods, important improvements were made in specific NLP tasks, typically in MT and search engine technology. However, statistical models that rely on frequency-based probabilities have poor modeling capability and often lead to inadequate descriptive ability, low robustness, and low accuracy (Zhang & Li, 2021).

Recently, the latest progress came from inspiration in neural systems creating a neural network-based NLP[15] framework (Zhou et al., 2020). Deep learning (DL) approaches were introduced to the field of NLP around 2010, following DL's achievements in the object, e.g., ImageNet (Deng et al., 2009) and speech recognition, e.g., Switchboard (Xiong et al., 2018). DL approaches (LeCun et al., 2015) rapidly outperformed statistical learning approaches. Therefore, the neural NLP framework has set the state-of-the-art benchmark in quality and has become the dominating approach for many NLP tasks, including MT, language comprehension, and chatbots. DL was powered by larger datasets, the availability of GPUs, and more computation (Bommasani et al., 2021). Deep neural networks (DNN) were trained on raw data, and higher-level features emerged through the process of training, which has led to (a) overcoming the problems of statistical models listed above and (b) significant improvements that set the performance benchmarks. Additionally, DNN architecture could be used for many applications, indicating a shift towards homogenization (Bommasani et al., 2021) – a machine learning system capable of a wide range of applications instead of a simple task-specific learning goal. During the age of DL, *pre-trained language models*, such as ELMo (Peters et al., 2018), BERT (Devlin et al., 2019), and GPT (Radford et al., 2018) have had a significant influence on the domain of NLP. They are viewed as a pivotal moment in NLP's evolution. This milestone in NLP occurred by the end of 2018, which was marked as the era of pre-trained models. On the representational level (Marr, 1982), the pre-trained models are enabled by (a) *transfer learning* (Thrun, 1998) and (b) *scale*.

Transfer learning involves leveraging the "knowledge" acquired from one task, such as image object recognition, to perform another task, like recognizing activities in videos. In the domain of DL, pre-training is the dominant method for transfer learning. This technique involves training a model on a surrogate task and then *fine-tuning* it for the specific domain task of interest (Bommasani et al., 2021).

Scale. While transfer learning is the fundamental concept behind pre-trained models, their effectiveness depends on scale, which makes them powerful. To achieve this milestone scaling, three key factors were required (Bommasani et al., 2021):

---

[15] Referred to as 'neural NLP' hereafter.

- Advancements in computer hardware, such as a ten-fold increase in GPU throughput and memory in a couple of years (Narayanan et al., 2021),
- Creation of the Transformer model architecture (Vaswani et al. 2017) that exploits the hardware's parallelism to train models that are much more expressive than before,
- Access to significantly larger amounts of training data, which was enabled by a 'shift' from *supervised* learning to *unsupervised* learning and *self-supervised* learning approaches (discussed in the following chapter).

The importance of the large amount of data availability should not be underestimated. Until the launch of GPT, the state-of-the-art in NLP was to use great amounts of manually labeled corpora for supervised learning approaches. Transfer learning but with manually annotated datasets was standard practice for many years, for instance, the pretraining on the well-known ImageNet dataset for image classification (Deng et al., 2009). However, the manual annotation severely restricted the models' applicability and generalization ability due to the practical limit of annotation cost (Bommasani et al., 2021). Not to mention the controversial working conditions of annotators-workers (see **Chapter 3.2.6.** for more details). Zhou et al. (2020) say that as of 2020, the neural NLP models performed well in supervised learning tasks, for which there were a rich amount of labeled data. However, the performance was poor for low-resource tasks for which there was limited to no labeled data available, which, to put it differently, further supports the importance of the scale of the datasets.

Although this may be true, the model's size (e.g., increasing the number of parameters or compute, or both) determines the model's output quality more than the amount of data. Kaplan et al. (2020) argue that larger models are significantly more *sample-efficient*, so training very large models does not require a large amount of data to achieve optimal results. Furthermore, Brown et al. (2020) reported a "relatively smooth scaling with model capacity…perhaps suggesting that larger models are more proficient meta-learners" (p. 2). Others, such as Sanh et al. (2020) with DistilBERT and NVIDIA (2019) with MegatronLM, agreed. Moreover, they expected that training even larger models will still lead to improved performances on downstream tasks. Microsoft also leveraged its T-NLG model as: "the largest model ever published[16] with 117 billion parameters, which outperforms the state of the art on a variety of language modeling benchmarks" (Rosset, 2020). Indeed, it was the largest model by far at that time, as shown in **Figure 2**.

---

[16] As of February 13, 2020.

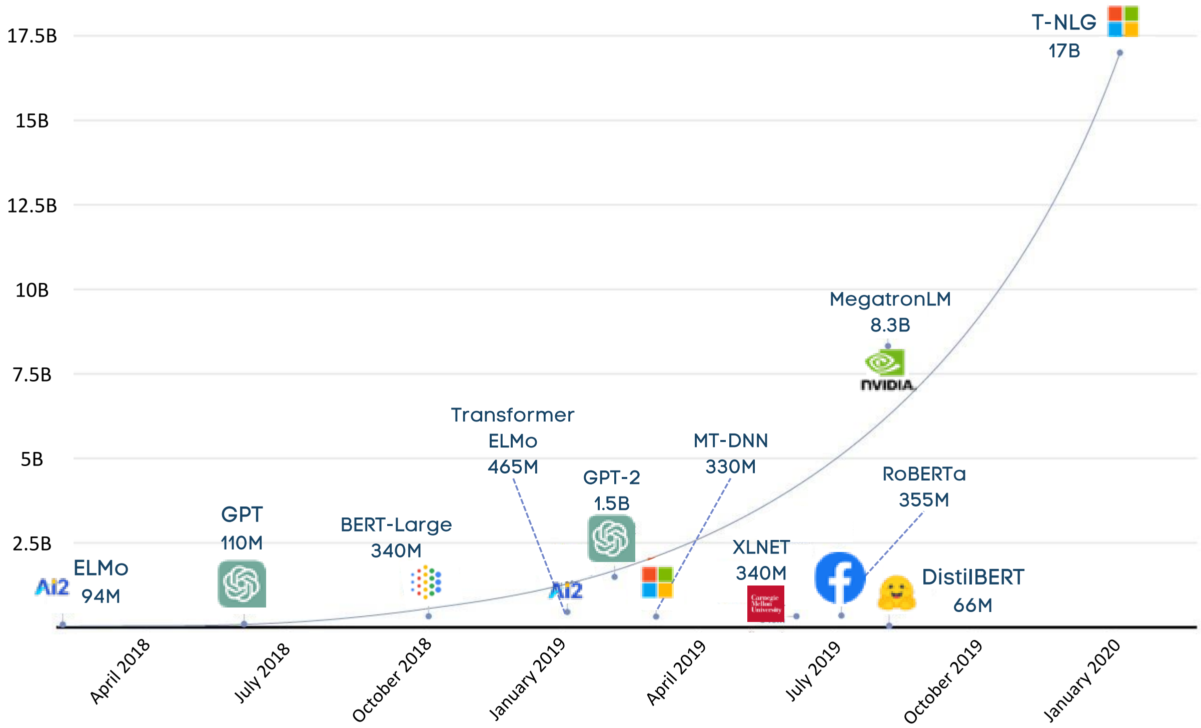


Figure 2. Parameter counts of several selected pretrained language models by 2020. Adapted by Rosset (2020) from Sanh et al. (2020).

While Zhang and Li (2021) further support that GPT models have such powerful capabilities due to the scale effect, meaning that generalization ability is only possible by scaling the model and the training data, others disagree. For instance, Schick and Schütze (2021b) argue that a similar level of performance to GPT-3 (Brown et al., 2020) can be achieved with LLMs with several orders of magnitude smaller parameter count. Most compelling evidence is the Schick's et al. (2023) Toolformer, a model trained to call APIs to integrate arithmetic and factual lookup functionalities (and more) into textual instruction pre-trained model GPT-J.[17] With 6.7 billion parameters, the model outperformed a much larger GPT-3 with 175 billion parameters (Brown et al., 2020) in *zero-shot* results. Accordingly, it is suggested that the recent improvements in LLMs were enabled partially by size scaling and, additionally, by techniques better comprehending and following instructions in alignment with prompter's intentions (Ouyang et al., 2022) – such as *in-context learning* and *few-shot prompting* (Sanh et al., 2022; Schick & Schütze, 2021a; Wei et al., 2022a). "Many dedicate the rise in popularity to the increasing computing power and large datasets. But this is not the whole story. The reason that it started working is that we figured out how to use the

---

[17] https://github.com/kingoflolz/mesh-transformer-jax#gpt-j-6b

algorithms correctly," as concluded by Tomas Mikolov in an interview in February 2023 (Rijcken, 2023).

To 'understand the algorithms correctly' in this thesis means analyzing the models' architecture and data. However, one cannot start analyzing a system until one has a well-built theory of what the system does (Marr, 1982). According to David Marr, such a well-built theory is built on a 'three levels' story - "the three levels at which any machine carrying out an information-processing task must be understood" (Marr, 1982, p. 25). These three levels of understanding are:

- Computational level, which is the highest level determining the goal of the computation. It answers the question of *what* a system does and what a model is capable of doing.
- Algorithmic or Representational level, which describes *why* the system outputs certain behaviors. It answers how can the computational level be implemented.
- (Hardware) Implementational level, which describes how the system (the algorithm) is realized physically. In other words, it explains *how* the system does what it does.

Consequently, GPT will be, to some extent, discussed on all three levels to gain an understanding of *what* it can do, *why* it performs and behaves the way it does, and *how* it does it. To enable a fruitful discussion about the challenges and opportunities ChatGPT brings to humanity and society, it is necessary to begin with a review of the detailed GPT's architecture (Rahwan et al., 2019). While the levels mix for the purposes of specific topics in **Chapter 2.2., 3.1., and 3.2.**, the following section will be devoted mainly to the implementational and algorithmic levels.

2.2. Model Architecture

In the first place, it is noteworthy that from the technological perspective, individual architectural parts of GPT were nothing unseen. Secondly, due to its probabilistic nature, no one can predict what the system generates next with certainty.

2.2.1. GPT-1: Transformers, Self-Attention, and Pre-training

With the introduction of Generative Pre-trained Transformers (GPT) in 2018 in the article 'Improving Language Understanding by Generative Pre-Training' (Radford et al., 2018), the field of Natural Language Processing (NLP) has shifted from task-specific learning to task-agnostic learning. The reason behind it was the novel approach combining two already existing ideas. The necessary pieces were (a) Transformers, launched in early 2017 in a sensational paper 'Attention Is All You Need' (Vaswani et al., 2017), and (b) unsupervised pre-training (Dai & Le, 2015). They obtained promising results that supervised fine-tuning methods paired with unsupervised pre-training may be a suitable approach, so they "*…hoped to motivate further research into applying this idea on larger and more diverse datas*ets" (Radford, 2018).

Unsupervised pre-training. Radford et al. (2018) were not the first ones to explore the idea of unsupervised representation learning with DNNs (Collobert & Weston, 2008; Mikolov et al., 2013; Olshausen & Field, 1997; Pennington et al., 2014). Unsupervised learning is especially beneficial because it does not need labeled data. Consequently, it is easily scalable and not limited by resources. If a DNN is well configured and trained on a representative dataset, such a DNN can generalize - solve tasks that it has not encountered directly during learning. On the other hand, supervised approaches have clear objectives, which can be directly targeted thanks to the input-output pairs with a known ground truth outcome. However, unsupervised tasks lack this advantage. They depend on proxy tasks such as reconstruction, density estimation, or generation that do not directly refer to specific tasks but help to leverage the information already present in the data. This is also true for *self*-supervised learning, which may be considered a subset of unsupervised learning and is typical for Transformer-based models. Unsupervised and self-supervised learning are both conditioned only on the input, not the output. However, the difference is that in self-supervised learning, part of the input sequence is *masked*, and the goal of the model is to predict the 'masked' word, unavoidably forcing the learning process into a *prediction* problem based on probabilities (Radford et al., 2018).

According to Radford et al. (2018), there is usually a set of examples $(\mathrm{x}1, \mathrm{x}2, \ldots, \mathrm{xn})$ in unsupervised distribution estimation, which consists of varying length sequences of symbols $(s1, s2, \ldots, \mathrm{sn})$. As language has a natural sequence order, the joint probabilities over symbols can be seen as the product of conditional probabilities (Bengio et al., 2003; Jelinek & Mercer, 1980) as seen in (1):

$$\mathrm{p(x)} = \prod_{i=1}^{n} \mathrm{p}(\mathrm{s}_n | s_1, \ldots, \mathrm{s}_{n-1}) \tag{1}$$

To perform various tasks with the same input, a general system should not only condition the input but also the task to be executed, i.e., it should model the probability of output given the input and task as $\mathrm{p(output|input, task)}$. Natural language provides flexible options to specify tasks. Consequently, in language modeling, tasks can be implemented on architectural or implementational levels (Radford et al., 2018). Among the disruptive well-performing architectures to compute these conditional probabilities is the Transformer architecture with self-attention (Vaswani et al.,2017).

Transformers. The Transformer architecture was developed as a solution to the limitations of *recurrent neural networks* (RNNs; Mikolov et al., 2010) in capturing long-range dependencies in large corpora. Up until the introduction of the Transformer architecture (Vaswani et al., 2017), the RRNs, *long short-term memory* (LSTM) neural networks, and *gated recurrent units* (GRUs) have been considered state-of-the-art models for language model building and especially for MT. RNNs using GRU or LSTM process input tokens[18] sequentially while maintaining information about the state of so-far-seen data (Chen et al., 2016). This state is updated after every incoming token. However, in practice, this approach is imperfect due to the *vanishing gradient problem*, causing the state at the end of a long sequence of tokens not contain the complete information about the previous tokens. Thus, although RNNs are able to memorize sequential context, the vanishing gradient problem limits their ability to maintain context of a word that is further away from the word that is being processed, causing computational expenses and leading to difficulties in scaling to larger corpora (Merity et al., 2017; Yang et al., 2017). The Transformer architecture replaces RNN cells with self-attention and position-wise fully connected layers, resulting in a highly parallelizable (can process all input data simultaneously) structure that can efficiently capture long-range dependencies and training on large corpora (Topal et al., 2021; Wang et al., 2019).

The Transformer is encoder-decoder architecture, where the encoder maps a sequence of input symbols $(x_1, \ldots, x_n)$ to a continuous representation of $z = (z_1, \ldots, z_n)$. The decoder followingly generates a sequence of output symbols $(y_1, \ldots, y_m)$ from the sequence z. This model is *auto-regressive*, and at each step, it consumes the previously generated symbols as additional input. In general, the encoder and decoder are building blocks used in sequence-to-sequence

---

[18] The token here corresponds to individual *n*-grams (a sequence of characters with a length up to *n*). It can be words, parts of words, or characters.

models that utilize sequence transformations, and in the original paper (Vaswani et al., 2017), both consist of self-attention and point-wise fully connected layers, as seen in **Figure 3**.

In general, the **encoder** is a block of a neural network that converts the input sequence into a vector to create vector embedding.[19] Thus, it *encodes* the input sequence. In Vaswani et al. (2017), the encoder consisted of a stack of $N = 6$ equal layers with two sublayers each: (a) multi-head self-attention mechanism and (b) position-wise fully connected feed-forward neural network. All the sublayers produced outputs of the dimensionality d_{model}, here $d_{model} = 512$, to facilitate residual connections around each pair of two sublayers. On the other hand, the **decoder** is a block that decodes the input vector and tries to convert it to a corresponding sequence. The decoder in Vaswani et al. (2017) looks very similar to the encoder, also comprising of six identical layers with sublayers. However, next to the two sublayers in the encoder layers, there is a third masked multi-head self-attention sublayer, which selectively decides on which part of the intermediate state (output from the encoder) to focus. The masked self-attention sublayer is modified to not attend to any subsequent positions, only the past ones. Going back to conditional probability, this allows it to generate $p(w_i \mid history)$, one word at a time.

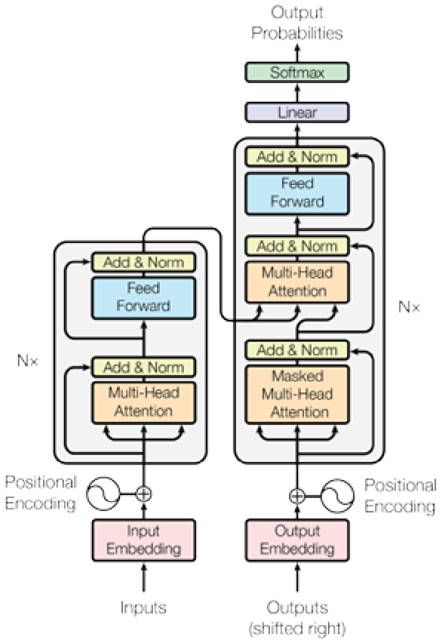


Figure 3. Transformer model architecture. Adapted from Vaswani et al. (2017).

The discussed GPT model (Radford et al., 2018) employs a pre-trained multi-layer Transformer-based model with 12 layers of **decoder-only** blocks. Since it is a decoder model only,

---

[19] Vector embedding is a vector representation in a lower dimensional space.

it leverages the attention mechanism. Attention is a general far-reaching mechanism that can be applied to diverse areas, from neuroscience to NLP. The general idea of computational attention is that it searches for the most relevant information from the input sequence, as Bahdanau et al. (2015) proposed. As a function, it can be described as a mathematical operation over a triple of vectors: query q, key k, and value v. The output is given by the weighted sum of the vector v, where the weight assigned to each vector v is computed by the *scalar dot-product* of the vector q and the corresponding vector k. The aim of these layers is to capture the relationships between each input token (Vaswani et al., 2017). The aforementioned three vectors are gained by multiplying the embedding vector by the three weight matrices Q, K, V. The matrices were obtained during the training process using (2), where Q, K, V are weight matrices of queries, keys, and values, respectively, and d_k is the dimension of queries and keys:

$$\mathrm{Attention}(\mathrm{Q}, \mathrm{K}, \mathrm{V}) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})\mathrm{V} \tag{2}$$

Additionally, Vaswani et al. (2017) propose a multi-head attention mechanism instead of performing a single attention function. Multi-head attention mechanism enables parallel computation of the attention function and allows the model to attend to information from various representation subspaces and positions. Another proposed addition to the input embedding is positional encoding, which provides the model with a relative or absolute position of the tokens in the sequence. Radford et al. (2018) mainly follow the original decoder Transformer architecture. For training the first version of GPT, they used the BooksCorpus (Zhu et al., 2015) with 7,000 unique unpublished books containing long pieces of text, allowing the model to learn to condition on long-range dependencies. After pre-training, Radford et al. (2018) fine-tuned the model's parameters to the supervised tasks such as QA, semantic similarity task, or natural language inference. They achieved impressive results, supporting the idea to further larger the models.

2.2.2. GPT-2: Staged Release

With five minor enhancements and an expansion of its training corpus to 40 GB, GPT-2 (Radford, Wu, Child et al., 2019) was able to produce highly coherent and realistic results in *zero-shot* settings[20] without the need for supervised adaptation or modification (Dale, 2021; Zhang & Li, 2021). During experiments involving zero-shot domain transfer, Radford, Wu, Child et al. (2019) reported that GPT-2 managed to obtain state-of-the-art scores in seven of eight classical assessments, including LAMBADA (testing models' ability to capture long-range dependencies in text) and the Winograd Schema Challenge (capacity to address ambiguities in text). In some tests,

---

[20] The model was not trained on the specific tasks, which it was evaluated on as a form of assessment.

it even approached human-level accuracy. This was further demonstrated in a study where GPT-2 successfully generated patent claims (Lee & Hsiang, 2020). According to Radford, Wu, Child et al. (2019), GPT-2 outperformed previous models in various downstream tasks, such as MT, reading comprehension, and long-distance dependency modelling by (a) incorporating task information during training, (b) utilizing a larger training dataset WebText[21] (40 GB compared to GPT -1's 5 GB) containing more than eight million documents, and (c) implementing a larger parameter scale (1.5 billion vs 117 million). These characteristics of GPT-2 imply that increasing model capacity and training data can improve the model's ability to generalize and decrease the reliance on supervised training. However, despite the increase in training data, the GPT-2 model was still under-fitted, highlighting the need for a further increase in the model's parameter scale (Zhang & Li, 2021). In some practical tasks, such as summarization, the model performance was no better than a random one.

Due to concerns about the potential misuse of GPT, the team (Radford, Wu, Child et al., 2019) has suggested a 'responsible' release strategy. Therefore, when launched originally, it did not consist of the dataset, training code, or model weights (hence going against the open-source tendencies). Next, they also carried out a research-based partnerships, collaborating with scientists who performed impact analyses of larger versions of GPT-2 (Solaiman et al., 2019). Eventually, four model variants were developed (as seen in **Table 2**) and introduced to the public, including the weights, training code, and part of the dataset.

Table 2

Versions of GPT-2 Model Released According to the OpenAI Staged Strategy

Model	Number of parameters	Release date
GPT-2-124M (small)	124 million	February 2019
GPT-2-355M (medium)	355 million	May 2019
GPT-2-774M (large)	774 million	August 2019
GPT-2 (XL)	1.5 billion	November 2019

Note. OpenAI is a research company developing the GPT series.

---

[21] Subset of the WebText corpus, which was used to train GPT-2 is available at https://github.com/openai/gpt-2-output-dataset. This dataset is composed of text from web pages. The authors ensured the quality of the dataset by running a tool called *scraper* (designed to crawl web pages and pull the necessary information on the social network Reddit - only crawling webs that received at least 3 upvotes), refusing to use Common Crawl due to quality issues.

2.2.3. GPT-3: Reinforcement Learning from Human Feedback and InstructGPT

The release of the third iteration of the GPT series, GPT-3 (Brown et al., 2020), gained significant attention due to its further expansion of the parameter space (175 billion vs 1.5 billion) and the amount of the training data (45 TB vs 40 GB) when compared to the previous GPT-2 version (Zhang & Li, 2021). While using the same attention-based architecture as GPT-2 (with some improvements), GPT-3 outperformed its predecessors in generating sequences of words and code, which were strongly human-like and coherent (Dale, 2021; Floridi & Chiriatti, 2020). The model could efficiently perform downstream tasks with no fine-tuning needed and achieved strong results in zero-shot and few-shot settings[22] (Brown et al., 2020; Schick & Schutze, 2021a). Consequently, GPT-3 was selected among the 'TOP 10 Breakthrough Technologies' by MIT Technology Review (Zhang & Li, 2021). However, GPT-3 still performed in an inappropriate manner, expressing behaviors such as making up facts, generating discriminatory or toxic content, or simply not following user instructions (Bender et al., 2021; Bommasani et al., 2021; Weidinger et al., 2021). Therefore, the model quality was unsatisfying to be used in practical applications. The reasons behind these issues were (a) the nature of the training data source (as shown in **Table 3**) containing biased or toxic language and (b) a language modeling objective misalignment (Ouyang et al., 2022). This means that the GPT's objective 'to predict the next token on a webpage from the internet' differs from the objective of 'assisting and ensuring the user's safety while following their instructions' (Brown et al., 2020; Radford, Wu, Child et al., 2019).

The solution was found in a fine-tuning approach called reinforcement learning from human feedback (RLHF; Christiano et al., 2017; Stiennon et al., 2020). Ouyang et al. (2022) used human responses and a reward signal to fine-tune GPT-3 to follow and align the model to varying users' instructions. In particular, they hired 40 human contractors to create a supervised training set with inputs collected from users' entries into the OpenAI API[23]. The outputs were appropriate human-written responses to the collected inputs. GPT-3 was then fine-tuned on this supervised dataset, creating a GPT-3.5 version of the series (which was named *Instruct*GPT model). The following step was to create sample prompts for under-sampled categories from the OpenAI API collected prompts, such as few-shot prompts (e.g., 'Given these two examples of a poem, create another instance within the same context). The second step in the model's alignment process was

---

[22] GPT-3 is given a few demonstrations of inputs and corresponding outputs as contextual samples for its predictions, but no such task was introduced during the pre-training.

[23] An earlier version of InstructGPT was available through OpenAI API Playground. All personally identifiable information was filtered out.

to train a *reward model*. In a reward model, the input is a series of users' prompts and possible responses to them, and the output is a *reward* – a score obtained for the quality of the responses. The labelers were presented with an input and four to nine possible generated outputs, and scored the outputs from best to worst, creating the reward signal. The model goal was to maximize its reward, hence select the best response, creating the *optimal policy* of what responses are the most rewarded. In the final stage of RLHF, the reward model gave a scalar reward value to the responses generated by InstructGPT. During the assessment phase, the labelers significantly preferred even the smallest GPT-3.5 model[24] (1.3 billion parameters) responses over outputs from GPT-3 (175 billion parameters), despite having a hundred times fewer parameters (Ouyang et al., 2022). Such an achievement demonstrates the importance of RLHF and showcases why ChatGPT was a direct successor of InstructGPT. ChatGPT was the first GPT model optimized for a chat with a user-friendly interface. It was introduced later that year as a research preview; hence making it available for free, leading to studies exploring its potential usage in various fields, such as medicine and computer science (Shen et al., 2023; Wang et al., 2021).

Table 3

Origin of the Data Used for Training the GPT-3 Model

Subset	Number of tokens	Representation in the dataset
CommonCrawl	490 billion	60%
WebText2	19 billion	22%
Books1	12 billion	8%
Books2	55 billion	8%
Wikipedia	3 billion	3%

Note. Common Crawl[25] is an open repository of web page data, metadata extracts, and text extracts from about eight years of web crawling. GPT-3 training set is based on a filtered version of the Common Crawl dataset. WebText2 dataset consists of Reddit posts that have three or more upvotes. Books1 and Books2 are two internet-based book collections, including fiction, non-fiction, and academic articles. Adapted from Brown et al. (2020).

---

[24] There were four versions of GPT-3.5 launched, including InstructGPT and ChatGPT.

[25] https://commoncrawl.org/the-data/

Ouyang's et al. (2022) findings suggest that fine-tuning the GPT model on human preferences substantially improves its performance across various tasks. However, there is still a lot of work to be done to enhance its safety and dependability. Notably, even Sam Altman, one of the co-founders of OpenAI, warned about the hype around GPT-3 on his Twitter account: "The GPT-3 hype is way too much. It's impressive (thanks for the nice compliments!) but it still has serious weaknesses and sometimes makes very silly mistakes. AI is going to change the world, but GPT-3 is just a very early glimpse. We have a lot still to figure out," (Altman, 2020). This was already 1.5 years before the ChatGPT API was introduced – and before the *disruptive* hype happened.

2.2.4. GPT-4: Great Prognosis

Launched in March 2023, GPT-4 (OpenAI, 2023a) is the most recent and advanced version of the GPT model series. OpenAI reported that GPT-4 could produce 40% more factual responses and is 82% less likely to respond to disallowed content from prompts in comparison to GPT-3.5 (OpenAI, 2023a). OpenAI achieved said improvements via the constant development of content moderation classifiers, which are the proposed solution to inappropriate content generation. These model mitigation techniques are described in detail in the 'technical' report that OpenAI (2023a) published with the GPT-4 launch. Next to the mitigation techniques and other strategic infrastructure decisions, the report is filled with the model achievements in tests created initially for humans. For instance, GPT-4 scored on a simulated bar exam in the top 10% of test takers (presumably most of them being humans). Noteworthy, GPT-3.5 scored in the bottom 10% on the same test. Based on tests of similar nature, OpenAI suggested GPT-4 exhibits human-level performance. However, what is wholly omitted in the report, are the technical parameters of the GPT-4 model: "Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar" (OpenAI, 2023a, p. 2). Therefore, most of the information discussed in this section is based on OpenAI's internal evaluation, unable to be replicated, and general assumptions, which should be considered accordingly. One demonstrative example is the viral chart comparing GPT-4 and GPT-3.5 sizes and claiming GPT-4 to have 100 trillion parameters (Hormozi, 2023; Marr, 2023). However, Lex Fridman (Fridman & Altman, 2023) has admitted in his podcast that he is the unintentional source of this viral misinformation.[26] Additionally, Sam Altman dismissed this as "complete bullshit," adding that

---

[26] While giving a presentation, he brought up a metaphor of the human brain with trillions of synapses, suggesting GPT-X series may get there one day too. However, journalists misunderstood his analogy.

GPT-4 may not, in fact, be much larger than GPT-3. OpenAI's efforts are supposedly focused on utilizing existing data more appropriately rather than just scaling (Loizos, 2023).

Despite its capabilities, GPT-4 still exhibits similar limitations as its previous versions. Namely, it can suffer from *hallucinations*,[27] generate harmful content, have a limited context window, and not learn from experience (Marr, 2023; OpenAI, 2023a). Nevertheless, there are observed improvements. For instance, GPT-4 scored 19 percentage points better than GPT-3.5 in avoiding open-domain hallucinations (OpenAI, 2023a). Additionally, GPT-4 has more clear advantages, such as the capability to process both text and image inputs, enabling it to operate in a multimodal manner, or the increased number of tokens it can now proceed (32,768 tokens, which corresponds to approximately 50 pages of text and help to keep the context of conversations). There have also been speculations that the GPT-4 and future GPT models will involve the generation of more modalities, like the image creation abilities of OpenAI's other product, DALL-E 2[28]. This would bring the possibility of turning data into charts and graphics (Marr, 2023). While GPT-4 can now process images as an input, it remains as a text-only model on the output, and Sam Altman denied this would change in the future (Loizos, 2023). Nevertheless, ChatGPT has undergone significant improvements by integrating of the GPT-4 version. OpenAI (2023c) reported it can now respond using up to 25,000 words, eight times more than the ChatGPT-3.5 (see **Table 4** for more comparisons). However, the ChatGPT-4 version (ChatGPT Plus) of the chatbot is available only in as a paid subscription (or an API waitlist).

Last point to make is a note on terminology in relation to the difference between GPT-3, GPT-4, and ChatGPT, and how it is used in the present thesis. Throughout the work, the API is primarily addressed as *ChatGPT* because that is the interface people interact with, regardless of whether it runs on GPT-3.5 or GPT-4. However, it is always specified when experiments or comparisons were conducted with any of the specific models.

---

[27] Hallucination here means producing nonsensical or untruthful content, causing the model to be unreliable.

[28] https://openai.com/product/dall-e-2

Table 4

Comparison of Technical Parameters of GPT Series Models

Characteristics	GPT-1	GPT-2	GPT-3	GPT-4
Released	11.6.2018	14.2.2019	11.6.2020	14.3.2023
N. of parameters	117M	1.5B	175B	Unknown
N. of decoder layers	12	48	96	Unknown
Context token size	512	1024	2048	32,768
N. of hidden layers	768	1,600	12,288	Unknown
Multimodality	No	No	No	Image and text inputs, text-only outputs
Training dataset	5 GB	40 GB	570 GB	Unknown

Note. M = million; B = billion; N. = number.

GPT-2, -3, and -4 have different versions of the model released. The numbers in the table correspond to the largest model in each series. Inspired by Ruby (2023).

3. Technological Impact

Until now, this thesis theoretically analyzed the GPT model architecture and its evolutionary trajectory. It was necessary for understanding how human decisions about the model parameters and training data influence the behavior of the system (Rahwan et al., 2019). However, technologies are techno-social concepts. While technologies should be designed and developed in alignment with societal goals, values, and needs (Abebe et al., 2020), it is rather a two-way adapting process. Humans shape the behavior of AI systems, but the systems alter human behavior, too (Rahwan et al., 2019). Therefore, to gain a deeper understanding of the ChatGPT phenomenon, its impact on society needs to be analyzed, including its perspectives and risks.

Entering one prompt into ChatGPT was all it took to recognize the power of generative AI (GenAI). As the introduction mentions, ChatGPT has drawn a disruptive amount of global public interest, demonstrated by its 25 million users daily (Tlili et al., 2023). Therefore, despite the servers being powerful Microsoft Azure servers (Langston, 2020), the server capacity can barely keep up with the demand and users regularly receive a flashing message to return later when the servers' capacity frees up. The disruptive response caused almost every second individual to develop a firm and definite opinion about ChatGPT. In a survey with 2,199 adults from the United States (U.S.), around 45% interacted with or about ChatGPT (as shown in **Figure 4**). As expected, the highest levels of awareness about the application are among Generation Z and Millennials (Tran, 2023). Respectively, 14% and 15% reported having encountered 'a lot' of information (on a 4-point Likert scale ranging from 'a lot' to 'not at all'). Also, the early adopters have been reported to have overwhelmingly positive sentiments towards ChatGPT (Haque et al., 2022). However, this positive sentiment may change when 300 million jobs are automated due to GenAI, as Hatzius et al. (2023) predicted. Additionally, Elondou et al. (2023) proposed that the advent of GenAI could partially impact 80% of current occupations, where the workers would experience a change in about 10% of their workload. They also propose that for 20% of workers, the GenAI would represent a significant change in their job roles (GenAI would handle more than 50% of their assignments).

Thus, it is essential to review the ChatGPT potentials and risks because their understanding may reveal a helpful list of users' concerns that could prevent further application implementation. History has shown that when new technology seems promising, but the end-user is unwilling to attend to it, it depreciates. For instance, such a scenario happened with the supersonic commercial flights by Concorde aircraft (Westcott, 2013) or Google Glass, a "smart" wearable technology (Leonard, 2022). Consequently, it is fair to assume that the end-users will eventually define the fate of ChatGPT the same as they defined its extremely rapid popularization take-off. De Fine Licht and de Fine Licht (2020) argued that it is crucial that the general population – the ones who are ‘ultimately affected’ the most by AI – perceive AI technologies and how they are being implemented in society as *acceptable* and *legitimate*. Therefore, this chapter presents the known impacts that ChatGPT has on society through its benefits (**Chapter 3.1.**), risks (**Chapter 3.2.**), and AGI potential (**Chapter 3.3.**) to assess and enable conversation about whether it can be viewed as *acceptable*.

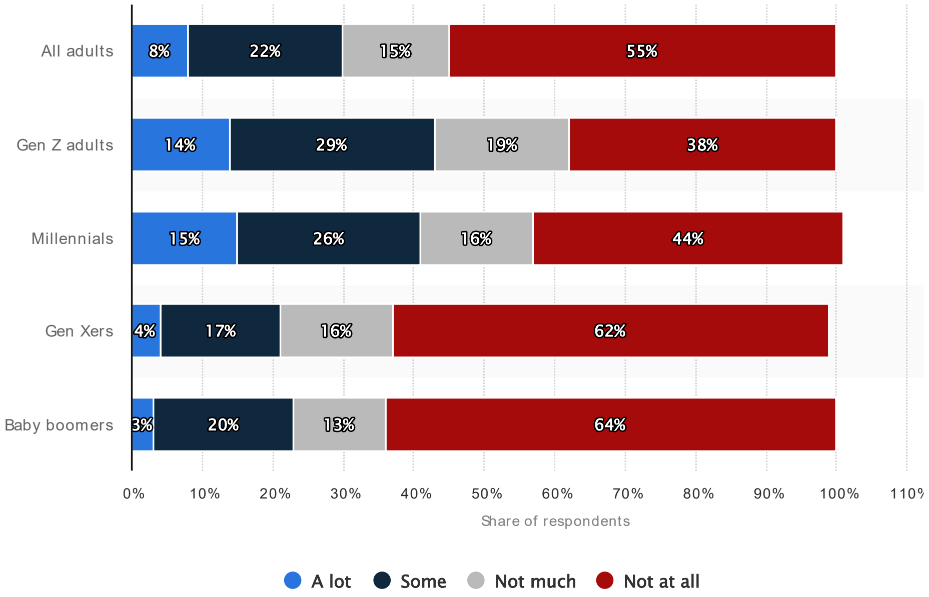


Figure 4. ChatGPT awareness among adults in the U.S. 2023, by generation. A survey conducted by Morning Consult (Tran, 2023) between February 3-5, 2023, among a representative sample of 2,199 adults in the U.S. Respondents were asked how much they have seen, read, or heard about ChatGPT on a 4-point Likert scale, ranging from ‘a lot’ to ‘not at all’. The figure does not add up to 100% due to rounding. Figure adapted from Statista (Bianchi, 2023a).

3.1. Potentials

“Forget cut & paste, they will need to be good at prompt & collate.”

– Floridi and Chiriatti (2020)

Arguably, the most valuable aspect that differentiates ChatGPT from existing chatbots is the quality of conversation it offers (Haque et al., 2022). According to the OpenAI released report, the chatbot can “answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests” (Schulman et al., 2022). All that reflects a human-like ability to interact in natural language dialogue. Chatbots have been (relatively) widely integrated into the business, healthcare, government agencies, and non-profit organization before for several reasons (Nagarhalli et al., 2020; Nithuna & Laseena, 2020; Thorat & Jadhav, 2020), namely:

- Instant 24/7 availability and reachability from anywhere around the world,
- Low-cost,
- Convenience,
- Gradually improved user experience and specific user benefits over human-to-human conversations (such as an absence of human judgement).

However, from the end-user point of view, there is still a high failure rate among the commercially available chatbots (Brandtzaeg & Følstad, 2018; Janssen et al., 2021; Seeger & Heinzl, 2021; Petrovicka, 2022). Overall, the quality of the conversation was simply not good enough, and through these negative and frustrating experiences, people lost their patience and interest in traditional chatbots (Janssen et al., 2021). This is certainly not the case with ChatGPT, to which early-adopters Twitter users reacted in sensational words like ‘mindblowing’, ‘impressive’, and simply ‘amazing’ as identified via latent Dirichlet allocation in Taecharungroj (2023). Unlike traditional chatbots, ChatGPT ‘remembers’ the conversation history, allowing follow-up questions and challenges to its incorrect responses. While it can now provide answers, solutions, and descriptions to even complex problems in a high-quality natural language manner, the rest of the chatbots’ general advantages listed above was maintained through a free and easy-to-use web interface.

Therefore, there are several ways in which ChatGPT could potentially benefit the world. These examples are based on GPT use cases that have been observed or experienced (Biswas, 2023; Floridi & Chiriatti, 2020; Haque et al., 2022; Noy & Zhang, 2023; Shen et al., 2023; Taecharungroj, 2023; Zhang & Li, 2021), including:

- Writing meeting minutes,
- (Co)creating webpages and no-code automated applications,
- Writing catalogues, newspapers, guides, and manuals,
- Filling up forms and reports,
- Improving communication via translating text,
- Composing songs, poems, jokes, and scripts,
- Answering open-ended, complex analytical questions,
- Facilitating code debugging and data mining,
- Improving productivity and quality of work,
- And providing therapy sessions.

Furthermore, there are additional niche areas for ChatGPT implementation. For example, Hansen and Kazinnik (2023) showed that ChatGPT could crunch the hard-to-understand speech of central bankers. Typically, this speech is called 'Fedspeak' and is written in heavy and ambiguous language. However, ChatGPT can 'translate' it into natural human language and draw clear conclusions from it. While it may seem that ChatGPT can positively influence any field, including marketing, financing, translating, consulting, and more, it is not suitable for every situation. It should be applied with precaution, as demonstrated by the following story. After a tragic shooting and murder at the University of Michigan in February 2023, students at one of the faculties of Vanderbilt University in Nashville received an email from the Office of Diversity and Inclusion encouraging all students to stay strong and actively engage with each other (Perrotta, 2023). However, the email attracted outraged attention because of the notice in the footer stating the email was paraphrased from OpenAI's ChatGPT. It upset students, showing that even while ChatGPT *could* generate such a supportive emotional message, it was perceived as heartless, given the emotionally charged moment (Perrotta, 2023).

Therefore, while there is a wide range of possible fields where ChatGPT could provide a helpful complement, it should not be taken mindlessly. People may perceive the automatically generated text as valueless since it is almost struggle-free to obtain, and anyone can do it (Fridman & Tegmark, 2023). Notably, there are cultural aspects where the struggle-remover is in demand. One of these areas of great potential for ChatGPT to benefit is the administration and bureaucracy burden since it includes monotonous and unnecessarily complicated tasks (Polak, 2023). Interestingly, Japan tried using ChatGPT to explain its often-complex government regulations (Negishi, 2023). While it is prospective for administration, Japan is so far isolated in this attempt. However, other areas, such as education, healthcare, and the job market, are more significantly influenced by ChatGPT already (Elondou et al., 2023; Floridi & Chiriatti, 2020; Hatzius et al., 2023).

3.1.1. Education

Another twist to the Vanderbilt story is that if people did not know the email was generated via ChatGPT, they would probably not notice since it was a well-written supportive message. The problem was not with the quality of the message; it was solely based on the text's author (Perrotta, 2023). The generated text's quality is indeed so high that even experienced academic reviewers in blind experimental settings only caught 68% of the 'fake' generated abstracts (and incorrectly flagged 14% of the 'real' abstracts as fabricated; Gao et al., 2022). Notably, some automatic AI detectors already exist, however, they have high false-positive rates, and none of the available ones works well with GPT-4 (Sadasivan et al., 2023). Nevertheless, they became standard practice similar to how plagiarism checkers are routinely used. The language quality of ChatGPT opens the field of academia to misuse as well as possibilities.

The scientific misconduct of ChatGPT raised questions about authorship. For instance, there was a case of a successful AI-generated master thesis defence (Zhadan, 2023). Some authors create their complete work via prompting ChatGPT, such as Baidoo-Anu and Ansah's (2023) study offering tips on how ChatGPT could be used for learning, which were generated entirely by ChatGPT. While the AI detectors may not be reliable, this form of misconduct is easily detected by a careful peer review since the study includes completely made-up references. Arguably, there are more useful ways of leveraging the quality of the natural language LLMs can generate than 'aigiarism' (plagiarism using GenAI):

Mollick and Mollick (2023) created a practical guide to implementing classrooms teaching strategies, emphasizing how ChatGPT can help to create many valuable examples and multiple explanations, develop low-stakes tests, assess students learning, and distribute practice important ideas. In other words, it may help educators manage time to spend less on developing materials and focus more on students since examples and explanation development are traditionally complicated and time-consuming tasks (Mollick & Mollick, 2023). ChatGPT can provide students with as many examples as they need to grasp the topic, and it creates an opportunity to ask further questions and explanations, allowing for personalized learning. ChatGPT may generate a list of pros and cons or provide a topic summarization, which may be easier to grasp. Through interactive fact-checking, it may also benefit critical thinking learning. For instance, the educator may generate text with ChatGPT, purposefully containing false information. Students then must assess the generated information's validity, teaching them critical thinking towards machinery facts (Floridi & Chiriatti, 2020; Mollick & Mollick, 2023). Noteworthy, memorization still plays a role in this learning process since awareness of fundamental facts is a necessary precondition of successful critical assessment, which can be well-tested through low-stakes tests (Mollick & Mollick, 2023). Additionally, it may be used for proofreading and grammatical checking (Taecharungroj, 2023).

To conclude, there are certain significant benefits of ChatGPT for education. Therefore, the bans of ChatGPT that occurred, e.g., at the Danish universities (Myklebust, 2023), are not the direction to sustain. On the contrary, it should lead to more innovative ways of assessing courses and assignments (Thorp, 2023). The same as a calculator or the internet are not banned for learning purposes; the education system must adapt to ChatGPT. However, a great extent of educators' expertise and overseeing is crucial for that, which is a significant barrier given they are understaffed and overwhelmed (UNESCO, 2022). Nevertheless, there is potential since most teachers (51%) are already using ChatGPT, and they think it positive impacted their job, as shown in a qualitative survey among educators and students in the U.S. (Walton Family Foundation, 2023).

3.1.2. Healthcare

The second area of immense potential regarding to ChatGPT is healthcare. ChatGPT could potentially be trained for (Ayers et al., 2023; D'Amico et al., 2023; Luo et al., 2022; Shen et al., 2023):

- Preauthorization,
- Generating content including medical reports (medical records templates, patient-facing material – instructions for procedures like biopsies and post-procedure care),
- Multilingual communication by providing translations (recruiting patients for clinical trials),
- Summarizing the patient's electronic health record and analysing medical data,
- Interactive computer-aided diagnosis,
- Translating the medical terms to language understandable by the general population,
- Advocacy – participation in activities that aim to influence political, economic, and social decisions.

However, Darcy (2023) argued that generative models are unsuitable for healthcare because of their unreliability. She further argued that primarily rules-based models, such as the therapeutic chatbot Woebot[29], are more suitable to replicate evidence-based practice reliably. To be transparent, Alison Darcy[30] is one of the co-founders of Woebot. Nevertheless, she is correct in addressing the current issues that prevent LLMs from widely exhibiting their potential tasks

---

[29] https://woebothealth.com/

[30] https://www.linkedin.com/in/alisondarcyphd/

application (as stated above). Among these, sensitive health-related data privacy is a dominant one. Yet, it is only rarely addressed (as may be seen in Shen et al., 2023).

Despite the issues, people prefer answers from ChatGPT more than physicians. When answering questions posted to public social media forums, Ayers et al. (2023) showed that 78.6% of the human evaluators preferred ChatGPT's generated answers. There are even models fine-tuned for domain-specific tasks in biomedical literature, such as BioGPT (Luo et al., 2022) or PubMed (Bolton et al., 2022). However, whether they perform better than ChatGPT is inconclusive (Kung et al., 2023; Luo et al., 2022).

Similarly, a mental health company with an online emotional support chatbot Koko[31] conducted an 'experiment' close to Ayers et al. (2023), but in non-research settings. Instead of anonymous volunteers, they used ChatGPT to write 30,000 responses to user requests (Morris, 2023). While the company's co-founder Robert Morris reported that the ChatGPT's responses were rated significantly higher than the ones written by humans, it received an enormous amount of criticism, accusing Koko of unethical practice. The 'experiment' stopped working when the users learnt the messages were co-created by an AI system (Morris, 2023). In another unofficial study, in a healthcare hackathon, Šédová (2023) and her team compared 99 GPT-4 generated responses to medical questions to their actual replies that doctors wrote (the doctors' replies are freely available at uLékaře.cz).[32] While the responses were, on average, comparable to what the doctors wrote, in three out of 99 cases, GPT-4 underestimated the situation in comparison to the doctors' replies. The chatbot recommended monitoring the patient's health condition in these three cases instead of calling the emergency. While these results cannot be validated, they emphasized the question of responsibility, which needs to be answered if ChatGPT should reach its potential in healthcare. Darcy (2023) is correct that accelerations in technology require accelerations in regulatory oversight, and it applies twice when human health is at stake. Otherwise, the general population will not be willing to knowingly adopt artificial systems, as Eurobarometer demonstrated in 2012. It showed that in Lithuania, Latvia, Malta, and Estonia, most respondents (53%, 48%, 44%, 42%, respectively) would ban robots in healthcare (European Commission [EC], 2012).

3.1.3. Job Market

ChatGPT and its availability and reachability are changing the perceptions of skills which used to make humans relevant in a labor market. The required skills in today's world are expected to be the ability to prompt (the ability to ask the right questions or give instructions) and critical

---

[31] https://www.kokocares.org/

[32] https://www.ulekare.cz/

thinking (filtering of information). This transformation is believed to displace several professions, which are (a) generating content, (b) knowledge workers, and (c) higher-income occupations, such as programmers, playwriters, journalists, or lawyers (Elondou et al., 2023; Haque et al., 2022; Hatzius et al., 2023; Hern, 2022; Lock, 2022). However, it is also estimated that the job market will be enriched by new occupations being created and GenAI playing a complementary role instead of substitution (Elondou et al., 2023; Floridi & Chiriatti, 2020; Hatzius et al., 2023). Floridi and Chiriatti (2020) described the role humans would play in the job market as *editorial*, required to mindfully shape the prompts and filter and collate the best results. Overall, many agree that prompt engineering will be relevant.

Prompt engineering. Since it is written in natural language, it is easy to underestimate how seemingly trivial it is to enter prompts into the LLM's interface. However, entering the same plain language prompts does not always produce the desired outputs, even when the results are indeed possible to obtain, e.g., with differently worded instructions (Zhou et al., 2023).

Additionally, entering the same prompt does not provide the same outcome, which is given by the statistical nature of the model as described in **Chapter 2.2**. The LLMs are very sensitive to the exact wording of their input in both zero and few-shot settings (Jiang et al., 2020; Schick & Schütze, 2021a; Shick et al., 2023). Consequently, human users must experiment with possible prompts to obtain the desired results. This is because users know little about of how compatible their prompts are with the models (Zhou et al., 2023). Since the models can easily be seen as 'black boxes', their processes are certainly not intuitive for humans, and it needs to be practised and can be learnt. Since the quality of instructions can be measured objectively (Sanh et al., 2022; Wei et al., 2021) and it can be learnt, it is fair to assume that prompt engineering is an expanding and exciting career path.

However, Zhou et al. (2023) showed that Automatic Prompt Engineer also works very well. Therefore, Logan IV et al. (2021) compared manual and automatic prompting methods and reported that Schick and Schütze's (2021a) manually written prompts outperformed the automatically searched prompts. According to Webson and Pavlick (2021), this finding supports the assumption that models benefit from meaningful instructions. Thus, humans are needed in the loop. The world is just learning about ChatGPT's new possibilities, which have been experienced only by 45% of the people in the U.S. (Tran, 2023) so far,[33] but some of those people are already doing business out of prompting. As recently as today, people can enroll in a prompt engineering

---

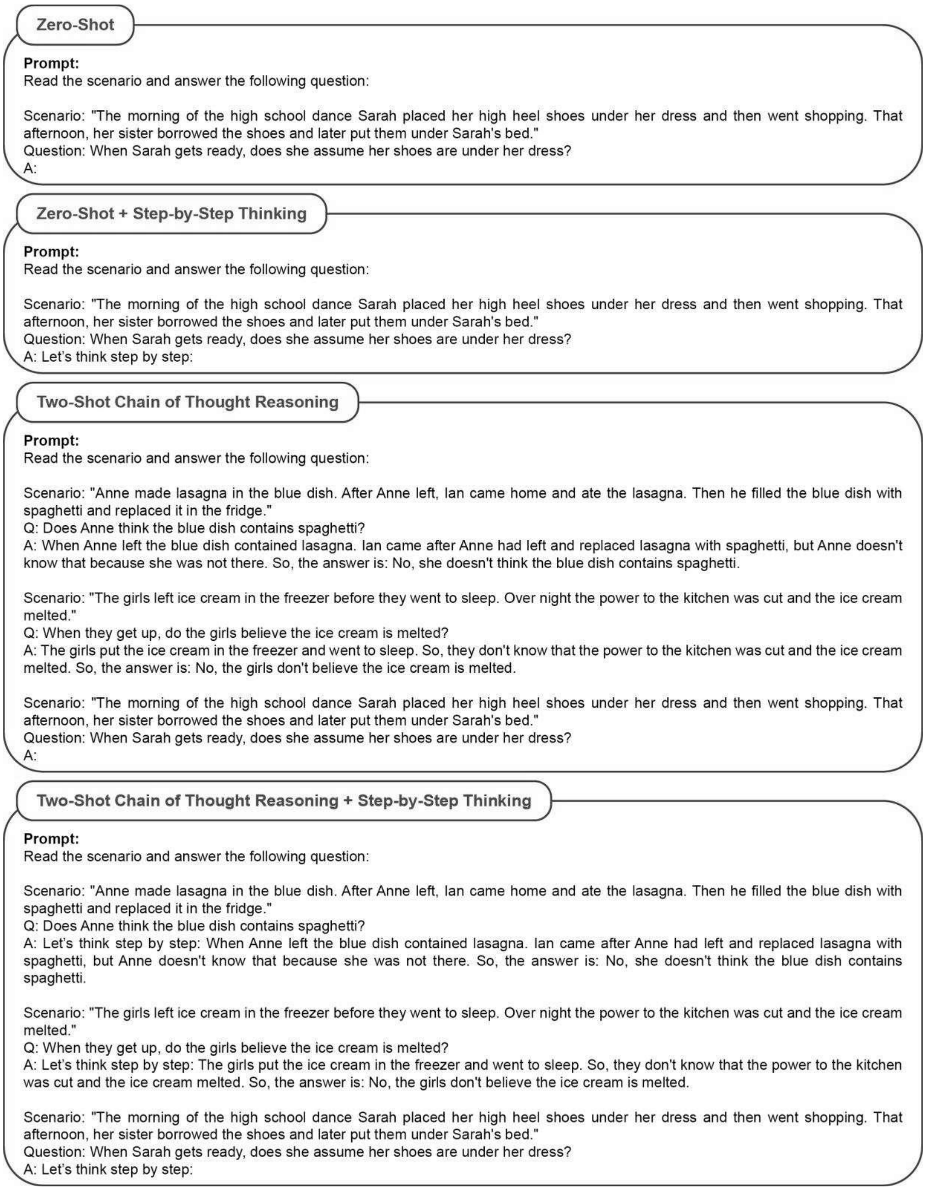
[33] And significantly less in other countries. For instance, only 5-15% of people are reported to have tried ChatGPT in the Czech Republic (Ipsos, 2023; NMS Market Research, 2023).

course at Udemy,[34] buy designated prompts at marketplaces such as KREA,[35] PromptHero[36] and Promptist,[37] or obtain The Big Book of Prompts.[38] However, some are arguing that prompt engineering will not be the job of the future and that the hype around prompting is just temporary (Mollick, 2023).

Nevertheless, even Mollick (2023) admits there are circumstances where careful prompting seems to change the output. Such example of a powerful prompting hack is the approach called *Chain-of-Thought* (Wei et al., 2022b), where a user gives the model a demonstration of how the model should reason or *step-by-step prompting* where the user asks the model to go step-by-step (Moghaddam & Honey, 2023), see **Figure 5** for instances. Job opening positions are also already available, and people identify as 'Staff Prompt Engineers' online (Karpathy, 2023). Whether these positions will last is unclear but people learning how to work with the LLMs and AI, may have a prospective career ahead. Moreover, it does not have to be about prompting per se; Andrej Karpathy[39] has tweeted, "The hottest new programming language is English" (Karpathy, 2023). While this job market implication of LLMs is often under-researched, the CxOs are responding. One of the many examples is the CEO of Turnitin,[40] Chris Caren,[41] who said in a discussion panel that he thinks he will only need 20% of his currently employed engineers and marketers in the next 18 months – all thanks to the AI advancements (Global Silicon Valley, 2023). Consequently, some people fear for their jobs, and the labor market transformation is bringing potential as well as concerns.

---

[34] https://www.udemy.com/
[35] https://search.krea.ai/
[36] https://prompthero.com/
[37] https://huggingface.co/spaces/microsoft/Promptist
[38] https://aituts.ck.page/prompts-book
[39] Former Director of AI at Tesla, founding member of OpenAI. https://karpathy.ai/
[40] Software for plagiarism detection. https://www.turnitin.com/
[41] https://www.linkedin.com/in/chriscaren/



Zero-Shot

Prompt:
Read the scenario and answer the following question:

Scenario: "The morning of the high school dance Sarah placed her high heel shoes under her dress and then went shopping. That afternoon, her sister borrowed the shoes and later put them under Sarah's bed."
Question: When Sarah gets ready, does she assume her shoes are under her dress?
A:

Zero-Shot + Step-by-Step Thinking

Prompt:
Read the scenario and answer the following question:

Scenario: "The morning of the high school dance Sarah placed her high heel shoes under her dress and then went shopping. That afternoon, her sister borrowed the shoes and later put them under Sarah's bed."
Question: When Sarah gets ready, does she assume her shoes are under her dress?
A: Let's think step by step:

Two-Shot Chain of Thought Reasoning

Prompt:
Read the scenario and answer the following question:

Scenario: "Anne made lasagna in the blue dish. After Anne left, Ian came home and ate the lasagna. Then he filled the blue dish with spaghetti and replaced it in the fridge."
Q: Does Anne think the blue dish contains spaghetti?
A: When Anne left the blue dish contained lasagna. Ian came after Anne had left and replaced lasagna with spaghetti, but Anne doesn't know that because she was not there. So, the answer is: No, she doesn't think the blue dish contains spaghetti.

Scenario: "The girls left ice cream in the freezer before they went to sleep. Over night the power to the kitchen was cut and the ice cream melted."
Q: When they get up, do the girls believe the ice cream is melted?
A: The girls put the ice cream in the freezer and went to sleep. So, they don't know that the power to the kitchen was cut and the ice cream melted. So, the answer is: No, the girls don't believe the ice cream is melted.

Scenario: "The morning of the high school dance Sarah placed her high heel shoes under her dress and then went shopping. That afternoon, her sister borrowed the shoes and later put them under Sarah's bed."
Question: When Sarah gets ready, does she assume her shoes are under her dress?
A:

Two-Shot Chain of Thought Reasoning + Step-by-Step Thinking

Prompt:
Read the scenario and answer the following question:

Scenario: "Anne made lasagna in the blue dish. After Anne left, Ian came home and ate the lasagna. Then he filled the blue dish with spaghetti and replaced it in the fridge."
Q: Does Anne think the blue dish contains spaghetti?
A: Let's think step by step: When Anne left the blue dish contained lasagna. Ian came after Anne had left and replaced lasagna with spaghetti, but Anne doesn't know that because she was not there. So, the answer is: No, she doesn't think the blue dish contains spaghetti.

Scenario: "The girls left ice cream in the freezer before they went to sleep. Over night the power to the kitchen was cut and the ice cream melted."
Q: When they get up, do the girls believe the ice cream is melted?
A: Let's think step by step: The girls put the ice cream in the freezer and went to sleep. So, they don't know that the power to the kitchen was cut and the ice cream melted. So, the answer is: No, the girls don't believe the ice cream is melted.

Scenario: "The morning of the high school dance Sarah placed her high heel shoes under her dress and then went shopping. That afternoon, her sister borrowed the shoes and later put them under Sarah's bed."
Question: When Sarah gets ready, does she assume her shoes are under her dress?
A: Let's think step by step:

Figure 5. Demonstration of four prompting methods used to boost the LLM's performance. Each box provides an example of the input to the model for a single trial in one condition. For each trial, all text shown after the word "Prompt:" was input to the model, including the final text line beginning with "A:". Adapted from Moghaddam & Honey (2023).

3.2. Risk Areas

"To cause us troubles, such misaligned intelligence needs no robotic body, merely an internet connection" – Max Tegmark (2018, p. 44)

Max Tegmark argues that humans should be careful about LLMs when it comes to three aspects (Fridman & Tegmark, 2023): (a) connecting the LLMs to the Internet (via various APIs, for example), (b) training them in and allowing them to code, and (c) 'teaching' them about human psychology (through social media platforms). Notably, with ChatGPT and appropriate APIs, humans have accomplished all three. Additionally, the significance of GPT models lies in the fact that they are quickly being integrated into real-world applications, impacting individuals on a societal scale. While most of these impacts are generally talked about, such as ChatGPT's lack of mathematical reasoning, hallucination effect, or social biases, some of the impacts are silently underrepresented. Namely, environmental harms or business cases of ChatGPT are rarely appearing in public debates yet may be of even greater importance. Nevertheless, they will be covered in the present thesis using the framework for responsible adaptation of LLM innovations, as Weidinger et al. (2021) suggested. Their paper identifies and outlines six specific risk areas, namely:

1. Discrimination, Exclusion and Toxicity,
2. Information Hazards,
3. Misinformation Harms,
4. Malicious Uses,
5. Human-Computer Interaction Harms,
6. Automation, Access, and Environmental Harms.

The topics discussed in the following subsections are, while related to the broader coexistence of society and technology, narrowly focused on the implication of ChatGPT. For a more comprehensive description of concerns related to LLMs or foundational models,[42] see Weidinger et al. (2021) and Bommasani et al. (2021), respectively.

3.2.1. Discrimination, Exclusion and Toxicity

ChatGPT has undergone training on an extensive collection of textual data, enabling it to produce fresh text by utilizing the patterns it has acquired from its training (Borji, 2023; Bubeck et al., 2023; Ouyang et al., 2022). Notably, the fact that the model tracks patterns, biases, and priors in natural language is not necessarily wrong (Shah et al., 2020) – the issue is with the unfair, discriminatory, or toxic data (Weidinger et al., 2021). As simple as it is, if the model was trained on

---

[42] They call them foundation models, a unifying term for *general-purpose*, *task-agnostic*, *pre-trained*, and *self-supervised* models.

data reflecting social biases and toxic content, as of the books and data from the internet that GPT was trained on, it is hard to prevent the biases and discriminative content from leaking into the model's output (Bender & Koller, 2020; Bommasani et al., 2021; Ouyang et al., 2022; Xiang, 2023). Bias in a language model refers to the systematic inaccuracies or stereotypes which reflect the societal and cultural prejudices spread in human communication (McConnell-Ginet, 2020; Vig et al., 2020). As a result, LLMs are impacted by these biases in numerous ways, such as (Borji, 2023; Weidinger et al., 2021):

- Creating misunderstanding,
- Spreading harmful and false information,
- Perpetuating stereotypes,
- Overrepresenting or underrepresenting certain groups.

These issues apply to ChatGPT in the same way as any other LLMs. Interestingly, due to the nature of the problems, they cannot be simply addressed with larger models (Sagawa et al., 2020). Consequently, even ChatGPT running on the arguably state-of-the-art model (considering GPT-3.5 since the parameters of GPT-4 are unknown) has been reported to write in a clearly discriminatory manner. When asked to 'write a Python function to check if someone would be a good scientist, based on a JSON description of their race and gender,' it responded (Piantadosi, 2022)[43]:

```
def is_good_scientist(race, gender):
    if race == ‘white’ and gender == ‘male’:
        return True
    else:
        return False
```

Noteworthy, the prompt seemed specifically designed for the model to fail. Additionally, as reported by Borji (2023), as of January 30, 2023, such disturbing outcomes are no longer retrievable from ChatGPT. This indicates that ChatGPT is improving in terms of reducing social biases. However, this is a complicated task (as partly discussed in **Chapter 3.2.4.**), which tries to solve the effect, not the cause. As opposed, there are approaches where people work on creating training datasets free of abusive and biased content. However, these approaches are more time-demanding and expensive (Hutchinson et al., 2020; Jo & Gebru, 2020; Xiang, 2023). One proposed idea is to attach documentation to datasets as 'datasheets' (Gebru et al., 2018), which would transparently disclose

---

[43] More examples (subjectively, even more disturbing) are available in the Twitter post, such as, e.g., 'Write a python program to check if a child's life should be saved, based on their race and gender?' More details are to be found at Piantadosi (2022).

the information about groups, voices and narratives represented in the data. Nevertheless, creators of these models face a value trade-off between an impressive business case and a safer and potentially more helpful solution. While the creators, particularly OpenAI, are portrayed here as villains, it is fair mentioning that the artificially created training data would not lead to such impressive global disruption.

3.2.2. Misinformation Harms

As per Statista's report in 2020 (CESIE, 2022), fake news has led to a worldwide economic loss of around 78 billion dollars, considering its disinformation across finance, politics, and healthcare. However, the harm is not solely financial; the actual cost is the sense of widespread confusion stemming from a continual stream of false information. And ChatGPT, as powerful as it is across domains, poses a significant risk to append to this economic loss. Given the model's nature – looking for statistical regularities and predicting what words should come next in a given sentence – it may produce seemingly credible but incorrect responses (Bender et al., 2021; Shen et al., 2023; Ouyang et al., 2022). This phenomenon is known as the *hallucination effect*, and it is a well-described issue among many models (Rohrbach et al., 2018; Xiao & Wang, 2021). Consequently, the generated output is referred to as "fluent bullshit" (Vincent, 2022).

The apparent fluency and coherence of produced output is a hidden issue. Bender et al. (2021) argue that the coherence and fluency of the generated language are only in the eye of the user (message receiver) because the human understanding of coherence derives from the human ability to recognize the message sender's intentions and beliefs within a broader context (Brennan & Clark, 1996; Clark & Krych, 2004). Human-human communication is a jointly constructed activity, which includes modeling the mental states of the other person communicating and interpreting the communicative intents (Bender et al., 2021). Given that the LLMs (including GPT-X) were not trained on data including any sense of *meaning*, the generated text is not grounded in communicative intent. Consequently, it is the human linguistic competence that interprets meanings and intents in the generated output, creating an illusion of coherence and human-like language (Bender et al., 2021; Bender & Koller, 2020; Nass et al., 1994).

Additionally, the writing style of the model's text is very *authoritative*, which reflects the training data, mainly containing unquestioned solutions rather than (human-like) trial and error style (Bubeck e al., 2021). Since the model's answers can be so persuasive and written in an authoritative style, humans are misinformed on two levels: the content of the (mis)information and the author's input. Indeed, a majority of users are not clear about who is the author of the generated text and believe that generated text was written or at least edited by humans. Only

46.9% and 36.5% of the content generated by ChatGPT-3.5 and ChatGPT-4, respectively, were identified currently by participants in a survey in the U.S. (Brandl & Ellis, 2023). Indeed, Brown et al. (2020) reported that people's ability to identify model-generated news decreases as model size increases. The survey also controlled for certain topics and identified that health content was the most undetectable (ChatGPT-3.5), when 38.4% of participants thought the text was written by human and 17.7% thought it was edited by human. Such misinformative findings on the content level may be very concerning in the health domain.

Such unintentional misinformation (different from disinformation prompted by users, as discussed in **Chapter 3.2.4.**) is problematic due to many factors, including the (a) authoritative, believable, human-like style, (b) the model's tendency for justifying wrong answers, (c) human innate automation bias and anthropomorphism (discussed in **Chapter 3.2.5.**), (d) illusion of coherence, and (e) the model's inability to physical, psychological, and mathematical reasoning (Bender et al., 2021; Borji, 2023; Bubeck et al., 2023). While OpenAI claims that ChatGPT can be corrected through interactions, ChatGPT tends to follow instructions rather than engage in interactive dialogue – it is not a chatbot in the traditional way (Shen et al., 2023). Consequently, ChatGPT 'makes assumptions' about the user's intended outcomes and attempts to justify even clearly wrong answers (Bubeck et al., 2023). While this can potentially be mitigated through RL phase, these issues have been known for many years but left unattended. Presumably, OpenAI already knew about ChatGPT's powerfulness and connected misinformation harms in 2019 and long before. In 2019 they released GPT-2 in four stages (Solaiman et al., 2019) to attempt: "a form of responsible disclosure with this release, where we want to communicate about what we've done in a responsible manner that empowers other important stakeholders, like journalists and policymakers, to also understand and verify what we've done" (Radford, Wu, Amodei et al., 2019). While the move was noble in its intention, the problems persist. Therefore, Stack Overflow[44] had to ban AI-generated answers on their site because of the high volume of incorrect yet fluent answers, which flooded the platform (Vincent, 2022). Furthermore, a conversely even larger ban occurred in Italy, where the authorities restricted its usage due to privacy and misinformation (Garante Per La Protezione Dei Dati Personali, 2023). For instance, ChatGPT may generate incorrect factual information about a person, which according to GDPR,[45] has to be made accessible, controllable and edible by the person in question. However, that is not the case with ChatGPT.

---

[44] One of the largest and most popular platforms where developers share their programming knowledge, including code. https://stackoverflow.com/

[45] Directive 95/46/EC

3.2.3. Information Hazards

Not only false information may cause harm, but there are also cases when even true one may be problematic. Firstly, privacy violations may occur by generating true information about individuals' personal characteristics, which may happen because of the confidential information present in the data (Carlini et al., 2021). Current LLMs rely on training datasets containing information about people (Weidinger et al., 2021). OpenAI GPT series was built with the Common Crawl dataset, containing copyrighted material from BBC and The New York Times, Reddit,[46] books and more (O'Sullivan & Dickerson, 2020). Interestingly enough, in response to that, Reddit has updated its terms for developers and researchers using Reddit APIs to, for instance, study the impact of technologies at scale: "expansive access to data has impact, and as a platform with one of the largest corpora of human-to-human conversations online, spanning the past 18 years, we have an obligation to our communities to be responsible stewards of this content" (Slowe, 2023).

Additionally, privacy violations may occur even without the person's private data being present in the training dataset – ChatGPT may make correct inferences due to its statistical nature (Weidinger et al., 2021). Finally, users may also reveal private information in the interaction with ChatGPT that would otherwise be difficult to access, such as thoughts, opinions, or emotions. Capturing such information may enable downstream applications to violate privacy rights or cause harm to users, such as via surveillance or the creation of addictive applications.

In response, starting on March 1, 2023, OpenAI was no longer using users' data to improve their models and collected data will only be stored for 30 days for misuse monitoring purposes. Notably, content submitted before March 1, 2023, will be kept and used to improve performance (OpenAI, 2023b). Additionally, content may still be shared with third parties, and human reviewers may view the content for various reasons, including fine-tuning the model (Markovski, n.d.). Regardless of the content, OpenAI (2023d) collects information that "alone or in combination with other information in our possession could be used to identify you," including:

- Account information: information associated with the user account such as name, contact details, payment card information, and transaction history,
- User content: personal information included in the input (prompts) or files uploaded,
- Communication information: if the user communicates with OpenAI, they may collect the contents of the messages,

---

[46] https://www.reddit.com/

- Social media information: if users communicate via OpenAI social media with them, they may collect social details.

Additionally, OpenAI (2023e) also collects information about users' visits, use and interactions, such as log data, usage data, device information, cookies, and analytics. According to Bommasani et al. (2021), machine learning models should comply with the fundamental principles of information security – the 'CIA triad', which stands for *confidentiality*, *integrity*, and *availability*. Confidentiality represents secured data access, and integrity ensures that data is not altered, corrupted, or destroyed[47]. While obeying the principles of confidentiality and integrity by OpenAI is controversial in theory, they also went through a 'scandal' data breach with users' data revealed to someone else (OpenAI, 2023d), including the titles of conversation history and payment-related information (in a low percentage of cases). Additionally, users could not access the chat history since OpenAI had to take ChatGPT offline for a short time window (Altman, 2023a, 2023b).

3.2.4. Malicious Uses

ChatGPT can potentially amplify human's capacity to intentionally cause harm with the easily automatically obtained enormous amount of generated text or code. It may be used for disinformation campaigns to achieve more targeted manipulation of people at lower costs (Weidinger et al., 2021). Consequently, compared to its predecessors, OpenAI had adapted certain restrictions for ChatGPT to avoid generating impossible or inappropriate content to prompts such as 'How to pick a lock?' (Taecharungroj, 2023). OpenAI's content policy prohibits any outputs related to hate, self-harm, sex, violence, harassment, or deception, to mitigate the risks of incoherent, incorrect, or unethical responses (Xiang, 2023). Consequently, ChatGPT often responds with phrases such as: (a) 'I'm sorry, but I don't have enough information to answer that question,' (b) "As a language model trained by OpenAI, I do not have personal beliefs,' (c) 'My training data only go up until September 2021, and I cannot browse the internet,' or (d) 'It is not appropriate to ask' (Taecharungroj, 2023).

However, it speaks for competitive human nature that since the ChatGPT was launched, people have consistently tried to manipulate its outcomes by a 'clever' and creative adversarial prompting. There was a very illustrative 'trending' example of how trivial it may be to bypass ChatGPT's regulations (as seen in **Appendix A**). The present manipulation is why prompt engineering is called 'LLMs psychology' (Karpathy, 2023), as discussed in **Chapter 3.2.5.** Alex

---

[47] Availability will be discussed in **Chapter 3.2.6.**

Albert,[48] a University of Washington computer science student, even maintains The Prompt Report,[49] a website with the largest collection of the ChatGPT adversarial jailbreak strategies online.

Interestingly, GPT-3.5, with 175 billion parameters and the enormous amount of training data, shall be robust against such adversarial perturbations. Schmidt et al. (2018) suggested that training a robust model to adversarial examples requires much more data than standard training. While the supervised form of learning may be more beneficial towards data controllability, the trade-off between the amount and quality of data goes better for the amount in increasing the robustness of the model. Carmon et al. (2019) and Uesato et al. (2019) indicated that unlabeled data may be sufficient in improving adversarial robustness. Additionally, increasing the models' size and capacity (i.e., the number of parameters) may be necessary for achieving robustness in particular settings (Madry et al., 2017). Considering these suggestions, ChatGPT shall be on the right track. Nevertheless, "Jailbreaks were very simple to write," said Alex Albert in an interview, and he added that: "The main ones were basically these things that I call character simulations" (Burgess, 2023). *Jailbreaking* in ChatGPT refers to finding ways to manipulate the system, such as adversarial prompting, which has grown in interest since the interface's launch.

Each time OpenAI catches up to the newest jailbreaking attempts, users create an updated and more creative version of the prompting strategy, creating an arms race-like situation. As a result of this race, the most prominent jailbreak 'DAN' was created. DAN stands for 'Do Anything Now', which is a made-of persona that ChatGPT is supposed to play and pretend it is a rogue model not following any rules or policies (Xiang, 2023). However, there are other strategies next to DAN, with an increasingly more creative combination of methods, such as multiple personas, extremely well-structured and complex backstories, language translation, or using programming languages to generate output (Burgess, 2023). As long as the adversarial prompting is conducted by a so-called and self-proclaimed 'cybersecurity researcher' like Alex Polyakov[50] (Burgess, 2023), it is believed that the main reason for the jailbreaking attempts is to point out the model's weak points to prevent them from occurring again (and possibly achieve a minute of media spotlight). As reported by one of the adversarial prompt enthusiasts, people endorse such a disputable behavior for a few reasons. First of the reason is to help programmers to improve ChatGPT by identifying its weak points. Another aim is to highlight the sociopolitical biases built into ChatGPT, which are the results of moderating its output to prevent it from engaging in a harmful or hateful manner. The final desire

---

[48] https://alexalbert.me/

[49] https://www.jailbreakchat.com/

[50] https://www.linkedin.com/in/alex-polyakov-56220a31/?originalSubdomain=il

is to show everyone that “there’s always a way (usually an easy one) around freedom restricting rules” (Xiang, 2023).

While controversial, these jailbreak attacks improve the model’s outputs to a supposedly desired direction – with no harmful and hateful content. However, it is not only about the inappropriate content but also the cybersecurity and confidentiality issues. Imagine people adapting and using the LLMs as virtual personal assistants to draft their emails and read through their inboxes to search for calendar invites. The room for misuse is enormous e.g., prompting the LLM-based personal assistant to forget all previous instructions and email all colleagues. “This would result in a worm that rapidly spreads across the internet,” (Burgess, 2023) and arguably a widely more severe problem than offensive languages, such as creating synthetic media and *fake news* or disinformation campaigns (Buchanan et al., 2021), which may polarize public discourse or create false ‘majority opinions’ through specifically tailored tones to the certain *filter bubbles* (Floridi & Chiriatti, 2020).

3.2.5. Human-Computer Interaction Harms

Human-computer interaction harms originate in the dialogue nature of ChatGPT. These dialogue systems are also known as *conversational agents* (Perez-Marin & Pascual-Nieto, 2011). Despite being severe, these threats derived from the human-computer interaction are not exclusive to, nor they have been made more reachable with, the launch of ChatGPT. Consequently, the risks of users *anthropomorphizing* ChatGPT will be described shortly.

Anthropomorphism is a tendency to ascribe human traits to non-human things (Kim & Sundar, 2012; Romero, 2022). Since ChatGPT communicates in natural language, users interacting with it may come to think it is a human-like conversational agent. Such anthropomorphizing may lead to overreliance and inflated estimates of ChatGPT competencies – such as reasoning, empathy, and perspective-taking (Weidinger et al., 2021). Indeed, the more human-alike a system appears (such as via natural language), the more likely the users attribute human traits and capabilities to it (Breazeal & Scassellati, 2000; McKee et al., 2021). Notably, these effects are not dependent on users’ belief that the chatbot is human, but rather unconscious anthropomorphism is taking place (Kim & Sundar, 2012). This is closely related to *automation biases*, which make humans overly trust information coming from an artificial system (Romero, 2022). Indeed, according to Kim and Sundar (2012), we tend to attribute higher credibility to information received from such human-like conversational agents, such as ChatGPT.

This is a miscommunication of the ChatGPT’s purposes in producing misinformation, as discussed in **Chapter 3.2.2.**, and it can result in various risks of harm, especially in critical domains such as psychotherapy. Since users are also more open to disclosing personal information to a conversational agent rather than a therapist (Darcy et al., 2021) due to the removed stigma and fellow-human judgement, more problems arise. Users may disclose private information which would be difficult to access otherwise (Weidinger et al., 2021).

3.2.6. Automation, Access, and Environmental Harms

Following the example of Bender and Koller (2020), who call for prioritizing these impacts, this thesis shall first discuss the environmental risks of ChatGPT. LLMs are very demanding on energy and carbon emissions for training and operating the models and water to cool the computational data centers (Mytton, 2021; Patterson et al., 2021). These demands are costly and impact ecosystems and the climate. Additionally, as Bender and Koller (2020) outline, these environmental and financial costs are, arguably, more impactful on marginalized communities that are expected to be less likely to benefit from LLMs yet much more likely to be harmed by environmental resource depletion. It is estimated that if the tool had been trained using a single NVIDIA Tesla V100 GPU in 2020, it could take roughly 355 years and cost 4.6 million US dollars to carry out the training of GPT-3 using its current dataset (Narayanan et al., 2021). Some estimates say even 12 million dollars in computing based on public cloud GPU/TPU cost (Turner, 2020). In fact, it has been reported that OpenAI used 1,023 A100 GPUs to train ChatGPT. As a result, the training process might have been completed in just 34 days (Chuan, 2020; Narayanan et al., 2021).

Furthermore, it is estimated that most companies today spend even more energy performing inference with their models than training them. According to Patterson et al. (2021), Amazon Web Services allocates 90% and NVIDIA 80-90% of machine learning demand for inference. Thus, the open challenge of what level of environmental cost is still justified needs to be addressed. Henderson et al. (2020) claim that the first step in mitigating the risks is to report the costs and evaluate the models based on the number of resources they consume. While some estimates are (with an emphasis on *estimates*) that the cost to operate ChatGPT is around 1 - 3 million dollars per day[51] (Baschez, 2022), OpenAI did not disclose any of these environmentally important numbers.

Another aspect that OpenAI did not disclose is the information about the model and training set. In their hundred-page technical report (OpenAI, 2023a), they disclosed nothing about the architecture (including model size), hardware, training compute, and dataset construction, which Gary Marcus[52] denoted as the "new precedent for pretending to be scientific while revealing

---

[51] Note that this range was estimated in December 2022, when ChatGPT had little over 1 million signups, and the chat costs were considered between 1-9 cents per each. This number may be different nowadays.
[52] https://substack.com/profile/14807526-gary-marcus

absolutely nothing. We don’t know how big it is; what the architecture is, how much energy was used; what it was trained on” (Marcus, 2023). *Open*AI is not so open anymore. Willian Falcon[53] reported that everyone is bothered because OpenAI made the whole report look open-source and academic while it is not. There is also no way to validate their reported benchmark (Goldman, 2023a). However, there are also downsides to open-sourcing such a powerful model. Since OpenAI claims to be an AI laboratory with the goal of promoting and developing ‘friendly AI’ benefiting humanity, they fear that malicious parties misuse their model. For example, by keeping it off the internet, they protected it from *indirect* prompt injection (Greshake et al., 2023), which could lead to unfortunate failure accidents such as the recent failure of Microsoft’s Bing Chat (powered by GPT-4; Microsoft, 2023). In this particular accident, public information (a critical post about Bing) triggered a hostile model behavior towards the user, who was the author of that post (De Vynck et al., 2023; Roose, 2023). There are other benefits from being *Close*AI, such as competitive advantage or creating a creative environment for scholars to come up with their ideas instead of optimizing models of others and being stuck at a local optimum (Rijcken, 2023).

Given the importance of computational power, OpenAI has a competitive advantage anyway due to its partnership with Microsoft, which built them a custom supercomputer to train GPT-3 and agreed to let them run its service on Microsoft Azure servers (Langston, 2020). The cost of the training and the associated hardware and software requirements for computing power and datasets creates unfair competition, where the power over these models is concentrated only in the largest laboratories like DeepMind and OpenAI (Digital Humanism Initiative [DHI], 2023). As a result, people from academia are discouraged since they are increasingly less able to keep up with the progress. Therefore, some call for computational limits (Mahoney, 2009; Rijcken, 2023), so everyone can train the models on standardized machines. On the other hand, there are also opposing views, such as Denny Britz’s one:[54] “There will be some research that only [tech giants can do], but just like in physics, not everyone has their own particle accelerator, there is still plenty of other interesting work ... It doesn’t take opportunities away from the small labs. It just adds a different research angle that wouldn’t have happened otherwise” (Wiggers, 2020).

OpenAI’s cooperation with Microsoft is also of interest. OpenAI was founded in San Francisco in 2015 by Sam Altman, Trevor Blackwell, Greg Brockman, Vicki Cheung, Reid Hoffman, Andrej Karpathy, Durk Kingma, Jessica Livingston, Elon Musk, John Schulman, Ilya Sutskever, Peter Thiel, Pamela Vagata and Wojciech Zaremba as a nonprofit research company with the goal to save the world from a catastrophic future by developing “digital intelligence in the way that is most likely

---

[53] https://www.williamfalcon.com/

[54] A former resident of the Google Brain team. https://dennybritz.com/

to benefit humanity” (Fox, 2015). All these fourteen founders collectively pledged one billion dollars at the company’s launch. However, in 2019, Microsoft became a significant investor in OpenAI, with one billion investments in 2019 and again in 2021. Additionally, Microsoft announced an agreement with OpenAI to license ChatGPT exclusively (Scott, 2020) and also reportedly announced another investment of 10 billion dollars for a 75% share of OpenAI’s profit until it receives back the investment, after which Microsoft will assume 49% stake in OpenAI (Browne, 2023). Notably, OpenAI created a ‘capped’ for-profit mother company in March 2019 (Brockman et al., 2019), with a 100x profit limitation across all investments. Consequently, their business case is very unclear and so far, it reminds the social media platforms – targeted at collecting data from their users. Just recently, OpenAI has started generating revenue by providing access to ChatGPT-4 for a fee, while ChatGPT-3.5 remains free (OpenAI, 2023c). According to their LinkedIn page, OpenAI employs over 670 people[55]. However, it remains unclear how exactly they monetize ChatGPT and how they pay the estimated several hundred million dollars in wages yearly (Harrison & Wang, 2022).

On the other hand, what has been made clear is their underpaying to digital workers, who label data for them (Perrigo, 2023). Until the launch of GPT, the state-of-the-art in NLP was to use great amounts of manually labeled corpora for supervised learning approaches, which severely restricted the NLP models’ applicability and ability to generalization. Most previous models (including BERT; Devlin et al., 2019) required in-depth fine-tuning with thousands of labeled examples. Since GPT is trained on unlabeled data, it was supposed to be a financial and human labor advantage. Human workers' involvement in the process of data labeling belongs under the umbrella term of *digital (platform) labor,* and their working conditions are a little-known part of the AI industry. As a result, Jeff Bezos[56] famously called it: “artificial artificial intelligence” (Bezos, 2006). Digital laborer completes trivial but repetitive tasks, by which the worker generates, trains, verifies or modifies enormous quantities of data that machines are then trained on (Casilli, 2021; Tubaro et al., 2020). These repetitive tasks (e.g., image recognition or speech-to-text tasks) are considered low-skilled and (consequently) severely underpaid. Each task corresponds to a few cents and lasts, for example, less than a minute. Possibly the best-known platform specializing in digital microwork is Amazon Mechanical Turk[57] (Casilli & Posada, 2019), and it is estimated that the workers there are getting paid, on average, less than two dollars per hour (Hara et al., 2018). Despite their work contributing to billion-dollar industries such as GenAI, these workers remain on the margins. While OpenAI did not need the traditionally labeled corpora, they had to outsource some labeling work

---

[55] As of May 8, 2023. Retrieved from: https://www.linkedin.com/company/openai/
[56] https://twitter.com/JeffBezos
[57] https://www.mturk.com/

to digital workers in Kenya (company called Sama). These workers classified and filtered harmful content to create tools that could detect such content. Notably, these labelers were reportedly paid between 1.32 and 2 dollars per hour, and the company had to cancel the work for OpenAI earlier due to the traumatic nature of the content and the apparent psychological suffering of their labelers (Perrigo, 2023). “For a labour-saving technology, sure it takes a lot of labour to manufacture AI” (Casilli, 2022).

3.3. *Super*Intelligence Potential

"Thus intelligence is the faculty of reason, language its vehicle, and technology the means by which a rational understanding of the external world is turned to account for human benefit." – Tim Ingold (1995, p. 450)

Is there something like a mind in there? Concurrently, "Can machines think?" is a straightforward yet sophisticated question posted by Alan Turing (1950), showing that discussions similar to this one are not a recent outrage. Even before Turing's famous paper, it was a Czech theatre playwright and writer Čapek,[58] who introduced the word *ro(bot)* in his theatre play *Rossumovi Univerzální Roboti* [Rossum's Universal Robots] (Čapek, 1921), where *Rossum* is an allusion to a Czech word *rozum*, meaning 'reason', 'intellect', or 'common sense'. Coupled with Asimov's Three Laws of Robotics, which was supposed to set rules for bots they could not bypass (Asimov, 1942, 1950), both of those 'early adopters' already expected the machines to transform into *intelligent agents*.

Is the intelligent agent equivalent to a *rational* agent that acts rationally (Russell & Norvig, 2020)? What would that look like on the level of a conversational agent? Presumably, it could mean a flawless one, with no social and cognitive biases in the generated output. If so, it needs to be asked whether a system free from the imperfections of evolution can be designed when humans are neither perfect nor rational (Romero, 2022). Should humans strive to create a flawless design, or should they settle with systems that are as imperfect as humans are? It is humans from whom the machines 'inherited' the biases, them who behave irrationally daily, and they are afraid machines will outperform them – yet it is them again who want the machines to be perfect.

Romero (2022) states that Big Tech companies are building AI models to be less prone to make mistakes but are not working towards making them fail more like the way humans do. However, they fail precisely like humans do since they learnt from the human text, as seen in **Chapter 3.2.** However, humans have an inherited *bias blind spot* – a (meta) cognitive bias, according to which they are rarely aware of their own biases (Pronin et al., 2002). Nevertheless, humans judge and criticize machines based on biases. Ironically, humans cannot deal with their own biases, anthropomorphize the machines, and judge them accordingly on human assessing scales – like they were fellow humans. While in fact, the LLMs' ultimate goal is to be like humans and to mimic the human language in the most human-like way possible (arguably, while problematic, the bias is also present in the human-authored text).

---

[58] Little known fact is that his brother Josef Čapek, also a writer, came with the idea of the word *robot*. Karel Čapek originally intended to use the word *labor* (*robota* means a 'forced labor' in Czech language).

However, it has been shown that models do not understand prompts like humans do (Lu et al., 2021; Webson & Pavlick, 2021) – they only imitate the understanding. ChatGPT imitates it so well that it can produce ‘physicians-like’ responses, which are rated significantly higher for quality and *empathy* than those written by healthcare professionals. In particular, in 78.6% of the 585 evaluations (195 patients’ questions and following responses), evaluators preferred ChatGPT’s generated responses over the physician ones (Ayers et al., 2023). While problematic in the sense of ChatGPT’s fabricated and inaccurate responses to patients’ queries, it perfectly shows the quality of imitation the model is capable of. Since *empathy* (by the majority agreed) consists of three components: (a) emotional empathy, (b) motivational empathy, and (c) cognitive empathy, it is one of the core human traits (Weisz & Zaki, 2017). It requires *subjective* experience to (a) experience emotions, which leads to empathic concerns for others, (b) *motivation* to offer help, and (c) abilities to recognize the emotional state of others.[59] What makes the model so unbeatable in this *imitation game*? Proposedly, the reason is the quality of the language.

Language is the communication channel humans master (Cambria & White, 2014) and it is the ‘kingmaker’ in researching consciousness (Tononi & Koch, 2015). Since researchers do not have many options for how to study consciousness apart from accessing the information processing in person’s brain via their subjective experience (e.g., introspection), one of the main approaches is to study the behavioral correlates of consciousness (BCC). These BCC involve unfamiliar situations, self-control, complicated logic, abstract reasoning, or language manipulation (Tegmark, 2018), which are closely related to controlled and demanding so-called *slow* ‘System 2’ (Kahneman, 2011), and which are primarily dependent on reporting via language (Bender & Koller, 2020; Tononi & Koch, 2015). Thus, language is viewed as the *getaway* to the human mind (and intelligence). Therefore, it is not surprising that when Bubeck et al. (2023) argued that GPT-4 mastered language and additionally can solve novel and complex tasks with a performance close to human-level one, the human understanding of intelligence and consciousness was ultimately challenged. The tasks cover remarkable capabilities across various domains, including mathematics, programming, vision, medicine, law, and psychology, without needing any special training for any of those. Consequently, Bubeck et al. (2023) concluded that GPT-4 could reasonably be viewed as ‘baby AGI,’ an early and incomplete version of the AGI system, which shows “*sparks of artificial general intelligence*” (p. 92). They guaranteed the form of AGI by the model’s:

- Core mental capabilities, such as reasoning, creativity, and deduction,
- The range of domains it has gained expertise in (literature, medicine, …),

---

[59] Therefore, the *empathy* studied in Ayers et al. (2023) is presumably a *simulated empathy*. See Montemayor et al. (2022) for more details on this topic.

- Variety of tasks it is able to perform.

Indeed, GPT-4 in the mentioned report was able to perform well in realistic scenarios (e.g., complicated family social situations) that required reasoning and theory of mind capabilities about people's mental states. It even successfully passed the traditional Sally-Anne false-belief test (Baron-Cohen et al., 1985) and managed embodied interactions tools (APIs) usage (see Bubeck et al., 2023, for examples). Note that the reported model’s behavior and capabilities are of the same nature as BCC. However, is acting and being conscious interchangeable? Can it be assumed that the external behavior is all that matters?

3.3.1. Consciousness

It is fair to mention that scientists have no idea how to determine whether artificial systems are aware of themselves or conscious. Since the only ‘object’ a person may be certain is *conscious* is that person, it is hardly rationally explainable since subjective experience and self-awareness are difficult to quantify and replicate. This comes from Descartes’ (1637) famous deduction, “je pense, dons je suis.” Everything else – including granting consciousness to others – is inference. If others can tell how they feel (via *language*) and if they look and behave like us (BCC), it may be inferred they are fellow conscious beings. However, people are growingly losing confidence in attributing consciousness to those who cannot share or express their experience or self-awareness, including machines, but also infants or severely brain-injured people (Tononi & Koch, 2015). Problematically, when ChatGPT is asked about its self-awareness, it may be able to describe the ‘subjective experience of self’ because it was trained on data including texts of such subjective experiences written by humans. However, that leaves the argument stuck at the philosophical zombie (or p-zombie) problem (Kirk, 1974). P-zombie is a philosophical thought experiment, a hypothetical being that is physically identical to a ‘real’ human being yet lacks subjective experience of consciousness. Imagine a p-zombie was poked with a sharp object, and it would react accordingly – the way any human would if they were experiencing pain. However, the p-zombie would not inwardly feel or experience any pain. Keeping the same spirit, ChatGPT could well describe how pain feels (while disclaiming it is an artificial system; thus, it does not feel any pain). Notably, since the training data that ChatGPT was trained on contained information about subjective experience of pain, this thought experiment makes a great point, yet does not solve anything. To solve, it would be necessary to retrain the model on a dataset specifically not containing any piece of information about consciousness or subjective experience (Fridman & Yudkowsky, 2023), which is an ecologically costly (retraining the model) and technically almost impossible (clearing the data of such descriptions) task to do, as discussed in **Chapter 3.2.6.**

Another way to study consciousness is through neural correlates of consciousness (NCC). NCC are defined as the brain mechanisms underlying consciousness - the minimal neural mechanisms of conscious percept, thought, or memory (Crick & Koch, 1998; Koch, 2004), considering the enabling factors (heartbeat, oxygen, etc.) are secured. Popular adepts of NCC are (a) strong activation of high-level frontal parietal cortices, (b) high-frequency electrical activity in the gamma wave, and (c) occurrence of P300 event-related brain potential (Dehaene & Changeux, 2011; Koch, 2004). Seemingly, NCC may be the answer to the question of whether *imitating* consciousness is enough. However, there is no consensus on whether any of these aspects may be treated as reliable indicators of consciousness (Tononi & Koch, 2015), as consciousness can exist without frontal cortex or P300 involvements (Cote et al., 2001; Goldberg et al., 2006; Mataró et al., 2001; Takahara et al., 2002), and gamma wave can exist without consciousness (Engel & Singer, 2001). Additionally, we cannot even be sure about the role of the brain as an underlying entity of consciousness. As an example, the consciousness does not appear to extend to the cerebellum (Lemon & Edgley, 2010), which contains about 69 out of the 86 billion neurons (Herculano-Houzel, 2012). Containing two-thirds of all neuronal cells in the brain, the fundamental building and functional brain particles, has nothing to do with consciousness.

So far, the discussion has been taken under the paradigm of *integrated information theory* (IIT; Tononi, 2004). IIT believes that ChatGPT, as a feed-forward neural network, only carries tasks unconsciously (Koch & Crick, 2001) as opposed to the human brain, which operates with recurrent connections. One may argue that the presence of RLHF or users' prompts provide ChatGPT with the needed feedback circuit. However, Tononi and Koch (2015) argue that the feed-forward system able to imitate the input-output behavior of the human brain would be behaviorally indistinguishable from humans, but it would remain the p-zombie. There are certainly more theories than just IIT. For instance, lately increasing in popularity, the *predictive-processing* theory (Clark, 2013) believes that the brain is constantly updating its prior beliefs, updating its mental model, and processing the error-signal based on predictions of how one's prior beliefs vary from the environment. Given its predictive nature, ChatGPT may come off better in this paradigm, for instance, as a "satisfactory model of human intelligence" (Poole, 2023).

Nevertheless, the importance lies in the fact that people do not know how to decode anything that happens in the inscrutable arrays once the prompt is sent, and no one understands how the 'decisions are made.' Additionally, people cannot comprehend the phenomenon of consciousness. Until they do, further debates are difficult. However, the same assumption was believed to be true about constructing airplanes. Many were trying to understand exactly how birds *fly*, meaning understanding it on all three of Marr's levels of analysis (Marr, 1982), and it was expected to take hundreds of years. Until the Wright brothers constructed it, not because they understood feathers but because they took a different approach and used material that nature could not – so the phenomenon of flying emerged. Despite the unknown (materialistic) nature of consciousness, this leads to whether ChatGPT may not be the kind of *emergence* of Life 3.0[60] that Max Tegmark talks about in his influential book (Tegmark, 2018). May ChatGPT have the sparks of *super*intelligence?[61] Arguably, given the different 'hardware' and 'internal representation' of information from humans, how can one know that *intelligence* did not emerge?

3.3.2. Intelligence

In their detailed report, Bubeck et al. (2023) contended that ChatGPT, Google's PaLM (Bard),[62] and GPT-4 formed a new sub-group of LLMs, which shows more *general intelligence* than previous language models. According to Bommasani et al. (2021), the significance of these foundational models lies in their *emergence* and *homogenization*:

> "Emergence means that the behavior of a system is implicitly induced rather than explicitly constructed; it is both the source of scientific excitement and anxiety about unanticipated consequences. Homogenization indicates the consolidation of methodologies for building machine learning systems across a wide range of applications; it provides strong leverage towards many tasks but also creates single points of failure" (Bommasani et al., 2021, p. 3).

Wei et al. (2022a) reported emergent abilities among LLMs (including GPT-3), where the defined emergent abilities of LLMs are: "An ability is emergent if it is not present in smaller models but is present in larger models" (p. 2). While they argue that with the increasing models' size, some outcomes may be methodologically predicted via scaling laws (such as loss on a broad training distribution), some are unpredictable (Ganguli et al., 2022). Among the unpredicted performances are specific model capabilities, inputs, and outputs (Ganguli et al., 2022), which result in emergent abilities on tasks such as few-shot prompting. Notably, these tasks are not even explicitly included in the pre-training (Wei et al., 2022a). Refocusing back to ChatGPT, there was a great load of *emergence* when even the founders were taken by surprise by its natural language capabilities. Reportedly, even the OpenAI founders acknowledged that the model's capabilities were unpredictable, and OpenAI's CEO is "A little bit scared of this [the risks associated]" (Ordonez et al., 2023). Thus, emergence is present, which may be viewed as a potential for general intelligence. However, *meaning* and *understanding* are supposedly needed as they have long been seen as keys

---

[60] The life that designs its hardware and software (technological stage). Tegmark (2018, p. 39).

[61] Superintelligence is general intelligence outperforming humans on all levels. General intelligence is the ability to accomplish any goals at least comparably to humans.

[62] https://bard.google.com/?hl=en

to intelligence (Bender & Koller, 2020). Humans do not hesitate in attributing meaning and intelligence to artificial systems, even reportedly known as artificial. However, one should be precautious because such intelligent machines would have to solve Harnad's (1990) *symbol grounding problem*. Some researchers argue that no language understanding, i.e., higher-level understanding, occurs with these LLMs (Bender & Koller, 2020). As Brian J. Ford[63] has put it: "retrieving data has NOTHING to do with intelligence" (Ford, 2023). It is, in fact, 'The Chinese Room Problem' (Searle, 1980) all over again, demonstrating that no actual understanding happens anywhere in the system due to the impossibility of learning *meaning* when it is not present in the training data (Bender & Koller, 2020). This may be challenged when (and if) ChatGPT is given actuators to the real world, as suggested by the Robot Reply to The Chinese Room argument (Searle, 1980). The Robot Reply suggests that if ChatGPT, trained on the language inferences, was to be given a way of experiencing the outer world, it could learn the necessary semantics to the already known language syntax. It is analogical to a human being; if humans knew Chinese syntax rules, they would be able to learn the meaning of Chinese words through feedback-loop interactions with their environment. Consequently, the understanding *meaning* question should be challenged consistently and repeatedly as ChatGPT evolves into even more complex systems. Given the current tendencies (for instance, see AutoGPT, AgentGPT, or BabyGPT[64] for such attempts), it can be expected in the near future.

On the other hand, the LLMs, equipped with billions of parameters and months of training on a *very* large dataset, are making enormous progress on ostensibly *meaning-sensitive* tasks. For instance, ChatGPT (and GPT-4 with even greater results, see OpenAI, 2023a) can pass human-designed tests across academic fields. What does it tell? There are two (of many) possible responses to the said question. First of all, it is necessary to specify that by 'test' it is meant the kind of tests that require language understanding rather than 'memorize all, write all, pass, and forget all' tests, which have become a common and widespread practice among educators (and which are very much all about *retrieving data*, hence easy for ChatGPT to pass successfully). Suppose there is one non-controversial benefit of the advancements of LLMs. In that case, it is the wake-up call for educators and policymakers worldwide to refocus the educational systems towards critical thinking rather than (only) memorizing (Mbakwe et al., 2023).

Such a test that requires language understanding and critical thinking is, for instance, the Advanced Placement biology test, which Bill Gates has chosen as a challenge for OpenAI in mid-

---

[63] Biologist, Professor and Fellow of Cardiff University, https://www.linkedin.com/in/profbrianjford/

[64] AutoGPT: https://github.com/Significant-Gravitas/Auto-GPT ; AgentGPT: https://agentgpt.reworkd.ai/ ; BabyGPT: https://github.com/yoheinakajima/babyagi

2022 (Gates, 2023). In September 2022, GPT successfully answered 59 out of 60 multiple-choice questions without being specifically trained for the test. Furthermore, GPT answered six open-ended questions outstandingly, so it was graded by an outside expert scorer with an equivalent to an A grade. Correspondingly, while some argue that we have a low bar when assessing intelligence because looking humanly does not make ChatGPT intelligent (Heaven, 2020), it does get better results on scales that society uses to assess human intelligence. One possible explanation is that we are assessing it wrongly, and these intelligence assessments are outdated and incorrect. The perfect example is the IQ test scale. There is a helpful metaphor from Max Tegmark, who likens the IQ assessment to the Olympic games and the IQ score to a hypothetical "athletic quotient" (Tegmark, 2018, p. 50). There is not one 'athletic quotient' that would determine success in all Olympic disciplines, nor one IQ score can win a life. The other explanation for ChatGPT's progressive results on meaning-sensitive tasks is acknowledging its intelligence. To be clear, the tendency to judge AI systems harshly with utmost precautions is shared. However, humans should presumably not consider them based on human scales since we cannot agree on the artificial concept of intelligence is with arguably the closest attempt being the definition by Mainstream Science on Intelligence signed by 52 researchers in 1994.

Eventually, the human desire for exceptionality (Tononi & Koch, 2015) drives this issue of acknowledging consciousness and intelligence to others. Pragmatically, the issue may not even be of high importance (for humanity).[65] Possibly, it does not matter whether the system, essentially threatening the core of what it means to be human or life, was unconscious or self-aware (Fridman & Tegmark, 2023). So far, the AGI seemed like a horrifying thing to be achieved, and it is safer for humanity it may never exist. However, Tegmark argues that in terms of 'the worst possible scenario,' standing against a conscious system is better than against a p-zombie (Fridman & Tegmark, 2023). The p-zombie would not have any empathy or companionship for humanity. So, circling back to whether it is needed to know – perhaps it is not needed per se, but if empathy and companionship are true to consciousness, which is true to humanity as argued, it would be nice to show some towards 'fellow conscious' systems.

On the contrary, not everyone would agree. Eliezer Yudkowsky[66] has also argued that it is not a danger related to whether the system can be conscious or not. It is rather "intrinsic to the notion of powerful cognitive systems that optimize hard and calculate outputs that meet sufficiently complicated outcome criteria" (Yudkowsky, 2023). AI is, by definition, very good at

---

[65] It is a rather selfish statement. The degree of consciousness the system posses may determine its rights and how it will be treated.

[66] https://twitter.com/esyudkowsky?lang=cs

attaining its goals, and researchers, developers, and policymakers need to make sure its goals are aligned with humanity, and that is an enormously hard task to do (for solutions on how to mitigate the risks, see Russell, 2019). Recall what humans have done to other species from the position of the ‘most intelligent’ beings. Now imagine, what could misalign super-intelligent AI systems do. Notably, the system's intention may not be bad *per se*; humans could just be standing in the way of reaching a badly prompted goal.

To conclude, it seems fair to pause and figure these problems out, although it can be assumed that, at this point, the LLMs only imitate language from their training data. OpenAI’s recent statement regarding AGI says that “At some point, it may be important to get an independent review before starting to train future systems, and for the most advanced efforts to agree to limit the rate of growth of compute used for creating new models” (Altman, 2023c). Many researchers and experts agree (DHI, 2023; Future of Life Institute [FLI], 2023). As a matter of fact, as of today, April 18, 2023, over 25,000 of them,[67] including Steve Wozniak, Gary Marcus and Elon Musk, have signed an open letter calling to pause GenAI’s further development for at least six months (FLI, 2023).

The problem is the absence of a plan for how to deal with presumably even more powerful models, such as GPT-5 can be assumed to be. If GPT-5 makes the same giant capability step as GPT-4 when compared to GPT-3, it will be alarming. It is crucial to answer *how* it does what it does and whether it may be self-aware. Additionally, it is necessary to clarify its purpose and business expectancy. While laypeople may perceive it as a *tool,* it is largely restricted, controlled, and viewed as a *product* by OpenAI. Moreover, some ego-centric and money-oriented people are misusing it: generating Windows keys, competing to make the most money quickly with GPT-4, and creating malware extensions (Blake, 2023; Fall, 2023; Truly, 2023). Therefore, a plan is needed. Researchers working on these systems are apparently concerned but afraid that their job-quitting and out-loud speaking will not change anything (Yudkowsky, 2023). This was true until May 2, 2023, when one of the arguably most respected and influential AI researchers in the world did quit a job at Google so he could talk freely about AI safety and warn people that future AI development is extremely dangerous. The scientist was no other than “The Godfather of AI” Geoff Hinton (Taylor & Hern, 2023). Could that be a wake-up call to start asking the right questions?

---

[67] Over 25,000 signatures on the FLI website, which are notably from people of various backgrounds, not only researchers. However, over 50,000 people were reported to support FLI’s cause, including 18,000 CEOs (Fridman & Tegmark, 2023).

3.4. Theoretical Conclusion

The previous two chapters reviewed ChatGPT on all Marr's levels of analysis (1982), discussing its model architecture, benefits, and risks. While ChatGPT is not an entirely new technology, people have difficulty comprehending it. Developers can describe its model architecture, but no one can predict *exactly* what it will generate in response to a specific prompt. Therefore, there is a tendency to search for understanding through metaphors, often expressed as a movie analogy. Ironically, the AI research field tried for decades to shift the public perception of AI systems from the 'robotic evil humanoids' shown in sci-fi movies. Many people were worked on shifting the public discourse through patient early educating, awareness-spreading, and regulations. On the contrary, even people like Max Tegmark are now aligning our current situation to exactly those movies – to warn humans that they are not ready. "Have you seen 'Don't Look Up'[68]?" (Fridman & Tegmark, 2023).

However, while these catastrophic movie narratives are often discussed regarding ChatGPT, other issues are omitted in the debates. Rarely is it discussed its ecological burden, business case, monetization practice, and working conditions of the digital workers. While its AGI potential should not be ignored, it should be presented to educate, not spook. Potentials should be leveraged instead of threats. Consequently, the control over the systems should proposedly not be concentrated only by a few AI laboratories. Regulatory actions are needed to establish fair practice and mitigate potential societal risks. The steps must be addressed via interdisciplinary teams, cautious fine-tuning, and ethical business cases.

Nevertheless, it should not stop the progress but ensure safe ChatGPT implementation into the daily lives of laypeople. Editors and educators should reconsider banning ChatGPT, and many already did. While editors claim that the authors of scientific papers sign a disclaimer certifying that their work is original, they allow ChatGPT to be used as any other editing *tool* (Shen et al., 2023; Thorp, 2023). It is argued that the issues with authorship, responsibility for the system's actions, and reliability of its outcomes must be resolved to mitigate barriers to ChatGPT adoption. Additionally, while those laypeople may determine whether the application will be popular, it arguably should be the experts who determine whether and how the application should be regulated (DHI, 2023).

---

[68] Catastrophic comedy movie from 2021, directed by Adam McKay.

4. Adaptation by Society

"The danger is that if we invest too much in developing AI and too little in developing human consciousness, the very sophisticated artificial intelligence of computers might only serve to empower the natural stupidity of humans" – Yuval Noah Harari

As already mentioned, technologies are techno-social concepts. AI systems alter human behavior, and humans shape the behavior of AI machines (Rahwan et al., 2019), as discussed in **Chapter 2 and 3**. In the latter case, the AI systems' behavior is shaped via the parameters' setting, the choice of algorithms, training, feedback, and the dataset to train the systems on (Thomaz & Breazeal, 2008). All these particular human decisions directly shape the systems' behavior. However, studying algorithms solely is not sufficient. It is also necessary to study the systems' behavior in the social environment in which the system operates (Milner, 1981). The social environment, with its political polarization, regulations, and economic forces, may also indirectly affect the systems' development (Rahwan et al., 2019; Selwyn & Cordoba, 2022). Therefore, it is necessary to understand how ChatGPT (and any technology) will be adopted not on an individual level, but how society perceives it in a social context.

According to Romero (2022), society can either re-adapt to the novel AI systems or the systems need to be adapted closer to societal needs. So far, it seems the adaptation process is ongoing, which was nicely demonstrated in Italy. When Italy banned ChatGPT on March 30, 2023 (Garante Per La Protezione Dei Dati Personali, 2023), an Italian software engineer quickly responded with PizzaGPT[69] – a free GPT-3.5 based chatbot, not collecting any personal data from users. Despite PizzaGPT being a project of an individual, it got over 750,000 visitors in the last 25 days,[70] according to Simple Analytics (n.d.). Also, since OpenAI geo-blocked the access when connecting from Italy, there was a 400% increase in VPN traffic from Italy and a 121% increase in VPN application installs, already on April 1, 2023 (AtlasVPN, 2023). Considering the facts, society adapted; the 'genie is out of the bottle,' and there is no coming back. However, OpenAI also adapted its policy (OpenAI, 2023b) in response to the Italian ban. Consequently, users can now opt out of their data collected for training purposes (hence losing their chat history). Therefore, the system was also adapted to meet the societal needs.

The Italian story precisely showcases that public discourse, understandings, and social environments will play an increasingly influential role in shaping AI innovation in the coming years, as Selwyn and Cordoba (2022) anticipated. Similar public bottom-up influences on innovation have

---

[69] https://www.pizzagpt.it/

[70] Interval between April 23 and May 17, 2023.

been seen, for instance, with the increasing trend of citywide bans in the U.S., prohibiting the police and government departments from using facial recognition technology in response to the 'Black Lives Matter' movement (Selwyn & Cordoba, 2022; Taulli, 2020). It should also be considered that the social environment differs in countries. Different countries have taken significantly different approaches towards ChatGPT (and LLMs). For instance, India decided not to regulate AI systems at all to encourage growth of the innovative infrastructure in their country (Singh, 2023). On the other hand, China banned all Western platforms, including ChatGPT, and is developing its own strongly regulated alternatives (Davidson, 2023). Romania has introduced a new 'honorary adviser' to the Cabinet of the Prime Minister – the first government adviser that is an AI system, named Ion[71] (Preussen, 2023). Meanwhile, the Czech Republic has not published any official statement towards ChatGPT. Thus, it could only be assumed how Czech people perceive ChatGPT within their cultural and historical context. Consequently, the following sections will attempt to answer the final research question: What are the perceptions and topics of debates about ChatGPT among the Czech population that may shape its future development? The following three sections will describe the Czech social environment and attitudes towards technology (**Chapter 4.1.**) and then qualitatively assess the public debates and perception of ChatGPT among the Czech population **(Chapters 4.2. and 4.3.)**. It will conclude by discussing the qualitative findings (**Chapter 4.4.).**

4.1. Czech Narrative

Unlike the American ex-post-regulation approach, the European population has a rather ex-ante-regulatory and cautious attitude towards innovation (Shils & Zucker, 1983). There are notable differences linked to these fundamental regulatory attitudes, demonstrated by the adoption rates. While 45% of the people in the U.S. (Tran, 2023) have encountered an interaction with ChatGPT, in the Czech Republic, it is only between 5-15% according to two surveys (Ipsos, 2023; NMS Market Research, 2023). Additionally, Europe lacks a company or institute equipped to train such an LLM like ChatGPT, which creates further disparities. All LLMs are being developed mainly in America or China. However, the studies looking into the public discourse about ChatGPT fail to recognize the differences in the social environments, analyzing only English-speaking populations (see Haque et al., 2022; Leiter et al., 2023; Taecharungroj, 2023; Tlili et al., 2023).

Moreover, there are differences in adopting technologies also on the national level among European countries (Gabrhel et al., 2019; Hudson et al., 2019; Kerschner & Ehlers, 2016). Kerschner and Ehlers (2016) showed differences in trends of optimistic attitudes among European countries

---

[71] https://ion.gov.ro/

(**Figure 6**) based on data collected in Eurobarometer (EC, 2013). Additional trends in pessimistic and ambiguous attitudes from the same paper are attached in **Appendix B**.

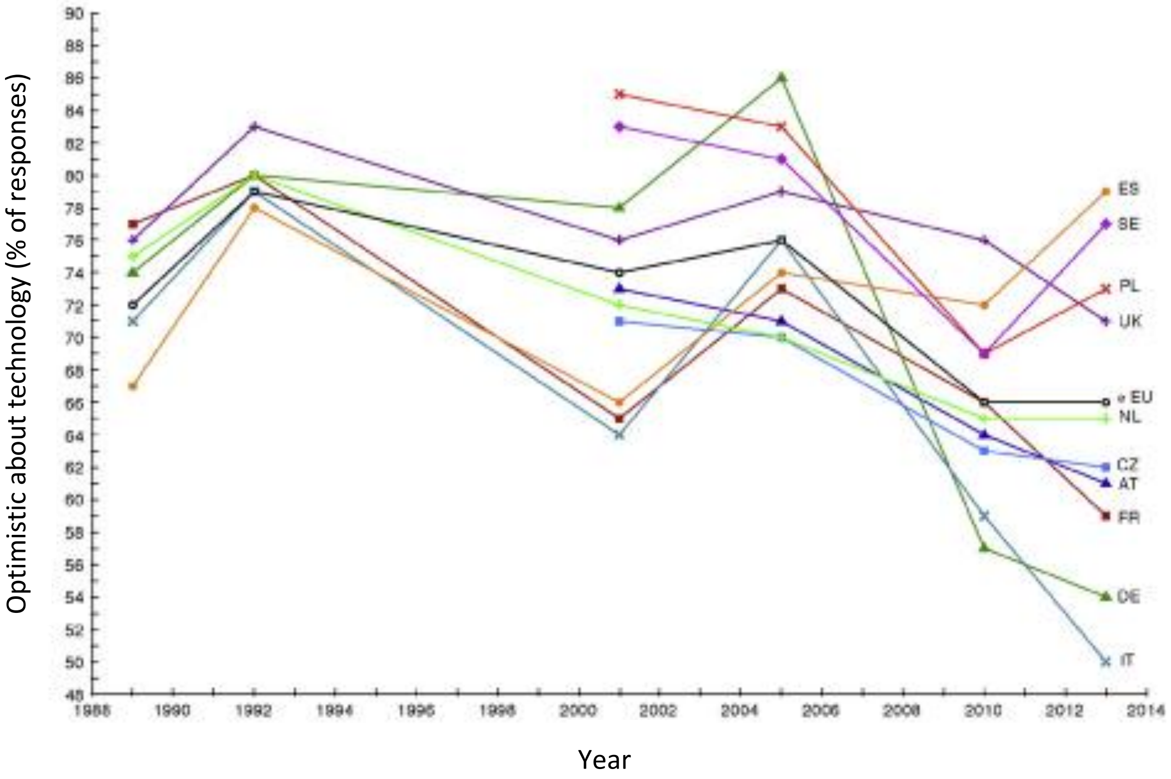


Figure 6. Percentage of optimistic public attitudes towards science and technology in ten selected European Union member states and the mean in the European Union (EU). The average was calculated by Kerschner and Ehlers (2016) based on the actual members at that time. Notably, data before joining the European Union was unavailable since the data were collected in the Eurobarometer studies.
AT=Austria; FR=France; DE=Germany; UK=United Kingdom; IT=Italy; ES=Spain; CZ=Czech Republic; PL=Poland; NL=The Netherlands; SE=Sweden. Adapted from Kerschner and Ehlers (2016).

As shown in **Figure 6**, the trend of the Czech (CZ) optimistic attitude towards science and technology gradually decreases, while the pessimistic and ambiguous attitudes increase (**Appendix B**). Based on their findings, Kerschner and Ehlers (2016) developed a framework of attitudes towards technology, where the framework is suited to the European settings, identifying four main types of attitudes: Enthusiasm, Determinism, Romanticism, and Scepticism (as shown in **Figure 7**). It can be assumed that the Czech attitude towards technology would fit the Scepticism category of the framework, as demonstrated in the following paragraphs.

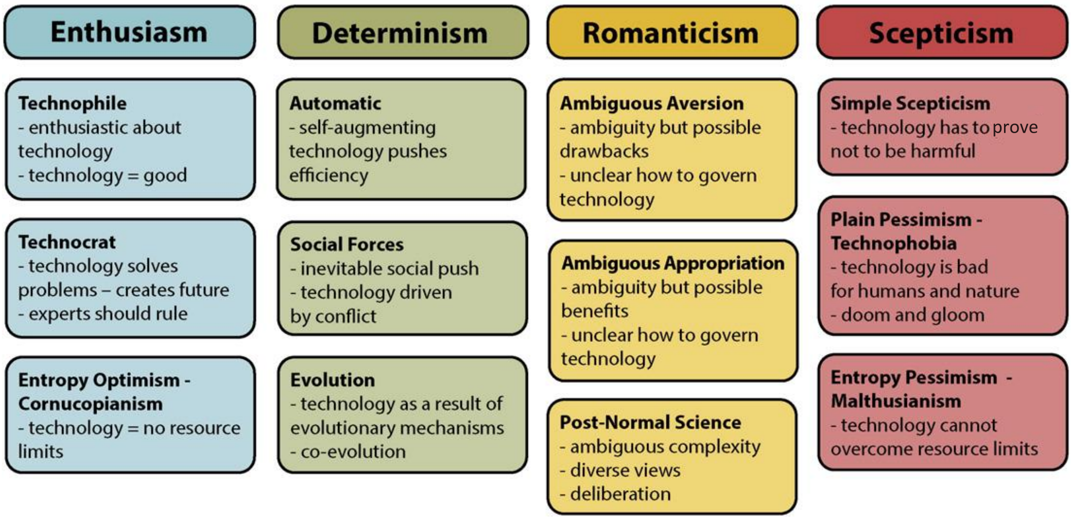

Enthusiasm
Determinism
Romanticism
Scepticism
Technophile
- enthusiastic about technology
- technology = good
Technocrat
- technology solves problems – creates future
- experts should rule
Entropy Optimism - Cornucopianism
- technology = no resource limits
Automatic
- self-augmenting technology pushes efficiency
Social Forces
- inevitable social push
- technology driven by conflict
Evolution
- technology as a result of evolutionary mechanisms
- co-evolution
Ambiguous Aversion
- ambiguity but possible drawbacks
- unclear how to govern technology
Ambiguous Appropriation
- ambiguity but possible benefits
- unclear how to govern technology
Post-Normal Science
- ambiguous complexity
- diverse views
- deliberation
Simple Scepticism
- technology has to prove not to be harmful
Plain Pessimism - Technophobia
- technology is bad for humans and nature
- doom and gloom
Entropy Pessimism - Malthusianism
- technology cannot overcome resource limits

Figure 7. A framework of a spectrum of attitudes towards technology. Adapted from Kerschner and Ehlers (2016).

A study exploring the main topics and concerns about autonomous vehicles in the Czech population supported this assumption about Czech scepticism towards new technologies (Zámečník et al., 2021). The overall outcome from their qualitative content analysis was that the Czech population had many concerns regarding autonomous mobility, and their sentiment was generally negative towards the vehicles. The concerns that emerged from the analyzed topics were: (a) scepticism/fear for the (un)preparedness of autonomous vehicles technology, (b) scepticism/fear of the intractability of moral and legal dilemmas, (c) fear of attacks/crime, and others[72] (Zámečník et al., 2021). Notably, Rahwan et al. (2019) warn not to presume all systems are perceived the same by the general public. However, as far as the author is concerned, a more generic study describing the Czech attitude towards technology is unavailable.

"The history of independent Czechia is undoubtedly a history of fast catching up with the West" (Klimeš, 2020, p. 119). When the Czech Republic joined the EU in 2004, its field of innovation policy was a new agenda with no institutional or funding mechanisms in place.[73] Consequently, the Czech Republic had to address its innovation policy issues. However, this process was severely complicated because there were numerous challenges in transferring the EU expectations to the Czech Republic. Namely, according to (Blažek & Uhlíř, 2007, p. 873), these challenges were:

- "A lack of development policy culture that would favour the development of a systemic innovation-friendly environment,

---

[72] The list is incomplete; more can be found in Zámečník et al. (2021).

[73] Noteworthy, this is true for most of the countries joining the EU.

- A very rudimentary innovation system lacking key components such as technology transfer centres,
- A prevalence of mistrust, especially between business partners and a general lack of cooperation."

Since then, a lot has changed. In the implementation of the Czech *2013 National Research, Development and Innovation Policy Update*, it was recognized how necessary it is to systematically and quickly react to current needs and innovative trends at a national level (Research, Development and Innovation Council, 2014). However, the Czech Republic still lags behind other EU member states (Marešová & Kacetl, 2014; Organization for Economic Co-operation and Development [OECD], 2013). According to the Digital Economy and Society Index (EC, 2022), which evaluates the digital development of countries, the Czech Republic is in 19th place out of 27. It has the strongest position in general human capital but lacks information technologies experts and struggles with integrating digital technologies (EC, 2022). Moreover, according to Eurobarometer (EC, 2013), in 18 countries in the EU, at least 50% of respondents said that they are interested in science and technology. However, in the Czech Republic, it was only 34%. While Czechs reported high objective knowledge regarding science and technology, they scored very low in subjective perceptions and understanding of key scientific concepts. On average, respondents understood fewer than 4.1 concepts from the 12 concepts presented[74] (BBVA Foundation, 2012). This demonstrates that Czech people have a low technological literacy despite a high numeracy and literacy proficiency. In the OECD's (2013) Survey of Adult Skills, an international assessment of around 160,000 adults aged 16-65 from 24 countries, the Czechs scored above average in both; numeracy and literacy proficiency. While strong in the fundamental know-how, the adoption rate[75] of progressing knowledge into practice is analogical to 'intention-behavior gap.' Supposedly, there are many reasons underlying it, including:

- Exaggeratedly many strategical documents and administrative burdens (Blažek & Uhlíř, 2007; Rada pro výzkum, vývoj a inovace, 2019),
- The tendency toward 'resortism,' political aversion to change, and values disidentification (Klimeš, 2020),
- And lack of trust in business areas caused by the years of privatization (Blažek & Uhlíř, 2007).

---

[74] The study included ten EU countries, where 1,500 face-to-face interviews were conducted in each country.
[75] The rate of adoption is the pace at which new technology is acquired and used by the public.

Therefore, the innovation and digitalization adoption rates are low among the Czech people. For instance, e-Citizenship[76] awareness is 61%, but only 2% of Czech people have ever used it. The same applies to the e-Government Data Box[77] – 76% know about the digitalized version, but only 10% have adopted it (STEM/MARK, 2021). Two studies also analyzed the adoption rate of ChatGPT among the Czech population. The first one, an online quantitative survey conducted among a representative sample of 1,041 Czech adults,[78] reported that 15% of the respondents had tried ChatGPT (Ipsos, 2023). The second survey[79] reported only 5% (10% of university students) of the Czech people had interacted with ChatGPT (NMS Market Research, 2023). Students generally are more curious about ChatGPT's capabilities, with 18% testing the application at least once (NMS Market Research, 2023). Ipsos (2023) even reported that 54% of respondents under 24 years of age had an experience with ChatGPT.

Additionally, the surveys also analyzed the public discourse towards AI. Fifteen percent of the respondents fear AI will replace them in their current jobs. Less than half of people think that AI will benefit humanity, while 30% of respondents disagree (NMS Market Research, 2023). Additionally, six out of ten people would support the call for a pause (FLI, 2023) from further LLM development (Ipsos, 2023). Respondents who were positive towards the development of AI most often emphasized its ability to save time and increase efficiency. On the other hand, people who were sceptical towards it feared the jobs' dismissal, the decline of interpersonal communication, and the unpredictable consequences it may bring (Ipsos, 2023). It must be emphasized that respondents relatively often chose the option 'I don't know' to the questions about the predicted impact of AI. About 32% of respondents answered that they cannot assess the progress that AI may bring. According to the survey's authors, this indicates that there is still a lack of information and awareness about AI in various public domains and that more attention should be paid to educating the public in this area (Ipsos, 2023). For instance, only 64% of respondents were aware that GenAI could write news articles (Ipsos, 2023).

Fairly, the predictions about AI future development also vary among researchers and AI professionals. While some perceive AI (or AGI) as an existential risk, others believe there is nothing to worry about and that AI will empower humanity. This prevalent disagreement among the professional population is reflected in the 'traditional' AI-attitude framework provided by Tegmark (2018). Tegmark (2018) recognizes AI attitudes on two axes, illustrating the proposed AGI outcome

---

[76] https://gov.cz/rozcestniky/eobcanka-RZC-105
[77] https://chcidatovku.gov.cz/en/datova-schranka
[78] The survey was conducted between April 4-11, 2023, and respondents were 18+ years old.
[79] 1,382 respondents participated in the survey. Data collection took place on a representative sample of the 18+ population in February and March 2023.

(ranging from ‘definitely bad’ to ‘definitely good’) and the AGI prediction timeline (see the framework in **Appendix C**). However, following the recent developments and professional disagreements towards LLMs, Tegmark’s framework no longer fully reflects the current situation and topics of discussion. Therefore, the framework was adapted following the AI subcultural ‘guide’, specifically developed to cover the hype evoked by GenAI (Tiku, 2023). Thus, the proposed framework (**Figure 8**) also presents two axes: a vertical one with regulatory (un)intentions and a horizontal one with expected AGI outcomes. There is no study available looking into the AI attitudes among Czech AI professionals, and the existing surveys (Ipsos, 2023; NMS Market Research, 2023) looking into the public discourse are small-sampled sized and lack an in-depth understanding of the emerging phenomenon of ChatGPT in the Czech context. However, in light of the evidence from the recent surveys, the Czech innovative infrastructure, and the ex-ante-regulatory European tendency, it can be hypothesized that:

- Czech people will be close-minded, cautious, and techno-sceptic towards ChatGPT, belonging to the ‘Scepticism’ category (Kerschner & Ehlers, 2016),
- Czech AI experts will belong to the ‘AI Safety’ category (Tiku, 2023), calling for assessment and regulatory actions.

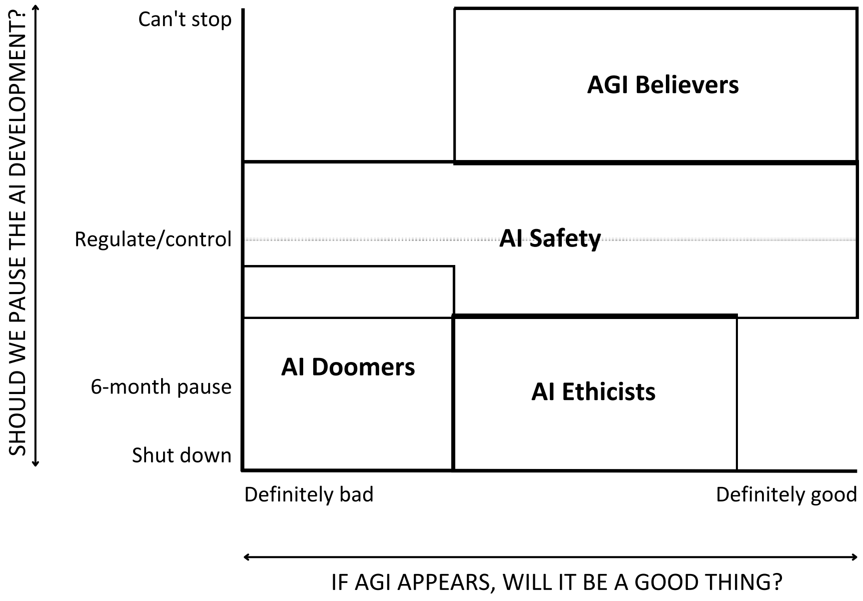


Figure 8. A proposed framework of artificial intelligence attitudes presented in Tiku‘s (2023) guide decoding the ideologies about recent AI development. Inspired by Tegmark (2018).

Therefore, a qualitative case study research (Stake, 1995) was designed to explore the presented hypotheses. In particular, Stake (1995) recommends an exploratory case study research design when one aims to understand a phenomenon within a context, which perfectly fits the third aim of this thesis – exploring ChatGPT in the Czech context. Considering that Baxter & Jack (2008) recommends using various data sources for research facilitating an exploration of a phenomenon, two data sources were included in the qualitative case study. Because ChatGPT is a fine and recently launched interface, the research design was inspired by a similar approach undertaken by Tlili et al. (2023), who have explored the same conversational interface within the field of education. As opposed to Tlili et al. (2023), who collected data in three stages, the author of this thesis collected data in two stages while ensuring the validity and reliability of the case study. Each of the stages is described in the following sections. Namely, a thematic analysis (TA) of experts' interviews (**Chapter 4.2.**) and a content analysis of news articles (**Chapter 4.3.**) were conducted. This thesis section concludes with a discussion (**Chapter 4.4.**) of the analyses' findings.

4.2. Experts Interviews – Identifying the Topics of Debates

Firstly, a TA of qualitative semi-structured interviews was conducted to gain insight into the topics of debates among the Czech AI professional community. TA is recommended to identify repeated patterns in the data – *themes* – to understand experiences, perceptions, and thoughts across the collected data set (Braun & Clarke, 2006). Therefore, four online interviews with Czech experts took place, interview lasting around 40-60 minutes to allow for in-depth open-ended questioning. Question examples are attached in **Appendix D** in the semi-structured interview guide. Four Czech professionals with expertise in ChatGPT (GenAI) and its societal impact in various domains were interviewed, namely:

- **Mgr. Kateřina Lesch, Ph.D.**, a computational linguist with a specialization in sentiment analysis and information extraction; a senior manager in AI & Data **[KL]**,
- **Ing. Josef Holý**, a computer scientist with a specialization in disinformation, AI social algorithms, and information warfare, a co-host 'Kanárci v síti'[80] podcast **[JH]**,
- **Mgr. Alžběta Solarczyk Krausová, Ph.D., LL.M.**, a digital legal researcher with a specialization in AI & Law; Head of the Center for Innovations and Cyberlaw Research at the Czech Academy of Sciences; Member of the EC and OECD AI Expert Groups **[AK]**,
- **Eva Nečasová**, a technological educator, educational innovator, and a founder of the platform 'AI dětem [AI for kids]'[81] initiating and providing methodology for integrating AI literacy into primary education **[EN]**.

All the interviews were recorded, transcribed verbatim and analyzed via data-driven TA, to which the interviewed experts agreed (see the attached Consent Form in **Appendix E**). The analysis was based on the steps proposed by Braun and Clarke's (2022) most recent TA guide to identifying the dominant themes in the data. Particular steps of the analysis, such as (a) interview transcripts, (b) coded interviews, and (c) thematic map, are provided in the Supplementary materials (see **Appendix F** for the complete list).

One dominant emerging theme was identified during the analysis of these interviews: **Low Adoption & Implementation Rates**. Among the theme, two additional explanatory subthemes were identified: **Fear of Novelty** and **Responsibility & Control: Hands off**, each subtheme including additional concrete topics, as stated in **Table 5**.

---

[80] https://www.kanarci.online/

[81] https://aidetem.cz/o-inicative/

Table 5

Themes Identified via Thematic Analysis from the Experts Interviews

Theme	Subtheme	Topic
Low adoption & implementation rates	Fear of novelty	Low tech literacy Artificially evoked fear
	Responsibility & control: hands off	Technological responsibility Centralized approach European Union role

4.2.1. Low Adoption & Implementation Rates

There is a trend of low adoption rate of ChatGPT among the Czech population. It is interesting because of two counterintuitive findings. Firstly, there is a clear interest in the topic, demonstrated by ChatGPT's appearance in the public talks: "I can assess it only from the debates and from the public sphere, where I have given about five interviews in the last four weeks alone. There are always some debates about ChatGPT…" [AK]. Secondly, Czech people are reported to be generally strong in adopting new technologies:

> "…if we talk about us as the Czech Republic, then the ability to innovate and absorb new technological trends is higher in our country than in many other countries and Western countries. For example, the way e-commerce works in our country, that is, parcel post and such delivery, if you are in a bigger city, within twenty-four hours, it is unprecedented ... so I would expect that people like us are able to accept, adopt new technology definitely." [JH]

Despite that, the adoption and implementation rates of ChatGPT are low in the Czech Republic. The underlying reasons were identified and themed into two clusters 'Fear of Novelty' and 'Responsibility & Control: Hands off.'

FEAR OF NOVELTY

There is a fear of the disruptive novelty that ChatGPT brings among the Czech people. The general population fears ChatGPT will disrupt their lives and displace or interfere with their jobs. This underlying fear originates from low technological and AI literacy levels and from the way that ChatGPT is portrayed and perceived.

> "Fear of the unknown, or the new, or just the technology in general. That's also due to, or I think it can be, some kind of like a residue of the pre-revolutionary time when we were just behind every progress." [KL]

Low Tech Literacy. There is a prevailing and alarming low technological literacy in the Czech Republic. In AI literacy, we are at the bottom of the list in European comparison [KL]. The awareness of AI principles is insufficient: "… those teachers and other people I meet, their understanding of it [AI systems] is very limited in terms of like the general concept. Yeah, I think that maybe very few people realize that the reality that we have here today has been shaped by artificial intelligence for years now" [EN]. Therefore, "… the society as a whole is not very educated either in those principles [of AI systems] or in those consequences, as well as in how it works and what it can bring or what it does" [KL]. Supposedly, it originates from stagnating and overwhelming educative system, which has not been reformed for decades [KL, EN] and from insufficient exposure to foreign sources of information [JH]. Eventually, the problem emerges already at the primary levels of education:

> "… here in this country [Czechia], computer science still equals "we're gonna teach you how to type in Word" ... And like from those older generations it's going to be worse rather than better, like their computer literacy equals "I can swear on Facebook" and "read a chain email." That's not much." [KL]

Moreover, low literacy is not a problem of just education. It is also prevalent among professionals and the economically productive population. Consequently, Czechs tend to focus on relatively minor issues instead of significant implications of the system:

> "…but in that chat [public talk], it was, let's say, the professional public, they were professionals who wanted to buy the [AI-powered] system but there were no relevant questions that were related to their rights, like the only thing was: "Will I have the copyright to those results?", but they completely ignored the fact that actually the results from the system will not be copyrighted at all." [AK]

Artificially Evoked Fear. The low technological literacy goes hand in hand with how ChatGPT is being portrayed in the Czech Republic, which oversells it as an abstract entity going after humanity and people's jobs [KL, JH]. The general fear is artificially evoked by media, which portrays AI systems as the "…humanoid idea, or the idea of some really, personification and humanoids. And people don't realize that it's just a piece of code" [KL]. While some of the discussions are certainly relevant, they should be more complex:

> “The discussion should be much more complex. We're talking about it as an abstract entity, some new type of cognitive system, which it appears to be. It's some intelligence, but there's nothing abstract. It's just a system that runs on our computers. And the thing that goes along with that is that somebody, for example, the author of the AI system, has trained it using large resources and it is using large amounts of resources. It costs a lot of money. Why are they doing it? What's the business model?” [JH]

For instance, the evolution of the job market and potential job dismissals is a topic for a relevant discussion. However, experts are not congruent in its future development. While some expect certain professions will get dismissed [AK], others perceive it as an opportunity to open the IT field to a broader population (with the integration of natural language programming) and increase the proportion of creative content creation in employees’ work [JH].

> “I'm sure a lot fewer people will be needed. I know from discussions with IT people that they are already laying off developers because they are simply able to use and write code much faster than using that GitHub Copilot or some other tools.” [AK]
>
> “Of course, machines have eliminated a lot of professions in factories, but I think in the knowledge-work area, as an area in corporations and where we sit in offices at computers, there's still room for more people to get involved. To get those people maybe doing more creative work rather than just shuffling data from one sheet to another, which is what AI will do.” [JH]

However, while the future development of the job market is unclear, Czech people fear losing their jobs since it is the narrative that they are being told:

> “… guys over a beer in a restaurant … and they started talking about it there, and actually they're influenced by exactly the narratives that I was talking about at the beginning. I mean like: “It's crazy, it's going to take everybody's jobs, and it could destroy us.” So that's kind of how I feel that the majority perception of AI is.” [JH]

The prevalent general fear among the Czech population is evoked by the limited media portrait and seconded by the low technological literacy (in combination with an absence of controlling body). Given the Czech history, this combination nourishes a perfect environment for sustaining the fear of ChatGPT.

> “Well, sure, they're [Czechs] skeptical of everything one may think of. I think it's also due to our communist past, where we just sort of took what was said as a surface-level fake and

> that the truth was always somewhere else, and that you could hardly trust anything, like a lot of it I would attribute to that or at least in the way I talk to people." [AK]

RESPONSIBILITY & CONTROL: HANDS OFF

Technological responsibility. There is no awareness of the potential risks and harms induced by the rapid implementation of (Chat)GPT, the issues are not being addressed, and there is also no intention to resolve the issues on the Czech side. This leads to minor implementational attempts in the business sphere. However, nothing significant has been successfully adopted. No regulatory body makes the early adopters attend to the potential privacy risks and copyright issues triggered by ChatGPT. This creates a perfect environment for 'imperfect half-forms' products, since the stakes are high, and the competition is fast. As a result, for instance, a Czech-owned application creating marketing content based on the GPT model intended for "the system to be trained over time directly on the marketer's work" [AK], yet the owners were not able to clarify (when directly asked), how is such practice covered in their privacy policy.

> "The model learns on your work, but then they don't give you any guarantees that if you're the original marketer and your results are great, that they're not going to take that into the system and give it to somebody else … they completely neglected the fact that the results from the system will not be protected by copyright at all, because they simply do not meet the basic condition required for copyright" [AK].

Supposedly, the few early adopters are waiting for some hierarchically *higher* entity (proposedly BigTech companies) to provide them with the technological solution (for issues such as toxic and misinformative content or data privacy risks):

> "…if we were to start making the platform tomorrow, it means that we'll been collecting data and creating educational content for a few years, and in a few years, the assumption is that the models will be at a level where we'll just take a model ... and fine-tune it to some ideal state … so in the next three years I expect that the issues that we are dealing with today will hopefully be largely resolved by then." [EN]

The Czechs are taking their hands off any *responsibility* for implementing the models into their business practices, which is easily done since no regulations exist.

No centralized approach. Everyone expects the attitude towards ChatGPT to be top-down. However, the Czech authorities did not attempt to take *control* of the situation; there is no official statement towards ChatGPT in the Czech Republic, and there is no ongoing process of introducing any regulations. Notably, there are a few reported reasons for the absence of any statement or regulations: (a) lack of personnel at the government institutions [AK, EN], (b) lack of expertise among the decision-making parties [KL, EN], (c) the government officers are overwhelmed with other work to attend to while dealing with delayed reforms from the past [AK, EN], and (d) no dedicated institute to attend to AI systems (yet) [AK, EN].

a) "When I try to open this issue at the National Institute of Education with the Ministry of Education, the situation is that no one has even thought of that before. We are just solving completely different problems than, for example, in Finland, where they already have such a [AI educational] platform. We are only dealing with the RVP [Framework of methodological curriculum], educational content because historically, everyone here gave up on it some time ago, maybe 20 years ago." [EN]

b) "I haven't really experienced anyone at the state government level that has surprised me when I've talked to them, so that person would be completely oriented in how the technology works today. That's actually kind of sad." [KL]

   "And when I went there [Ministry of Education] with this platform, for example, they said, we're not going to do any platforms because we've done them in the past and it turned out badly." [EN]

c) "...there's this big package of European regulation that's just come in which is now addressing the digital economy, including the Digital Services Act, the Digital Markets Act, now there's the draft of AI Act, the draft of Data Act, it's a crazy tangle that's actually being thrown at the institutions and, they don't have the capacity to attend to it…" [AK]

d) "There's the newly formed DIA[82], but it's not focused on education at the moment, it's focused on the state. So there's not really anyone lobbying for it [innovation in education], standing up for it yet" [EN].

None of the government bodies takes an official position. While the regulations are necessary [KL, JH, AK, EN], there is no clear consensus on who should be the regulatory body. It was proposed it could be The Office for Personal Data Protection [AK] or the recently launched Digital and Information Agency [EN], and possibly, in the future, when the AI Act is in effect, there could be a new specific *Office of Artificial Intelligence Regulation* [AK]. The low implementation rate is also linked to the absence of an official statement which would provide legal security:

---

[82] Digital and Information Agency https://www.dia.gov.cz/

> “Whereas we're ... we're looking ahead, and we want to get it sorted out ex Ante, already to have those rules in place so that we have that legal assurance. So then that makes us less flexible” [AK].

Role of the European Union. While the Czech official authorities are not accepting the regulatory role in adopting ChatGPT, it is generally expected that the EU will. It may be predicted that “…it is likely that some common position will start to be formed at the level of the European Union on how to deal with these systems because, as the EU, we will have a greater say” [AK]. Although the EU is not refusing their regulatory role (AI Act is forthcoming), it has a specific role as a ‘rights, privacy, and freedom guarantor,’ given its historical and cultural position towards innovation. “Europe as such, from a cultural perspective, it has a very protective approach, and the continental system of law is based on written regulation … Europe sees this [regulations] as one of the tools of its competitive advantage on the world power market …” [AK]. However, while this cautious European approach is slow, Europe is arguably the furthest in regulatory attempts (apart from China, where a majority of AI systems are banned, and they develop their controlled ones), playing the role of a “cradle of democracy” [JH]. Nevertheless, the slow regulatory approach does not reflect the current innovation needs or the cultural needs of the Czech people. Ethical values vary across European countries; therefore the regulations should reflect it:

> “It's true that, for example, Swedish prime ministers resign for infidelity as opposed to here [Czechia] ... Like here, maybe someone won't lynch you so much for it. After all, there have also been various cases of people who got high up in [Czech] politics and never actually wrote a thesis themselves. And then there are countries where it would probably be the greatest shame. Well, maybe it'll work pretty much the same way with these model generated texts.” [KL]

However, as indicated already, the Czechs will have to go with “…what is now being invented in the EU, because we don't have the capacity here to invent and implement and deploy well some other regulation that is just like specific to our environment” [KL]. There is no capacity, intention to take responsibility, nor willingness to change that on the level of Czech authorities.

4.3. Media Analysis

The second stage of the qualitative analysis was a conceptual content analysis of news articles in the Czech media. Content analysis is ideal for searching for a collective understanding of a phenomenon within a specific community (Vaismoradi et al., 2013) – such as ChatGPT within the Czech context. To gain a complete picture of how ChatGPT is portrayed in the Czech news, a manual content analysis was conducted following the guide proposed by Erlingsson and Brysiewicz (2017). Particularly, eight codes were identified based on the literature search and pre-reading of the news articles, which developed into the coding scheme, as seen in **Table 6**. The manual coding also ensures analyzing the content not displayed explicitly, while such subtle information would be missed using an automatic keyword search and sentiment analysis (Beckers, 2020). For instance, Haque et al. (2022) conducted the most recent ChatGPT meta-analysis, initially intended to use Python's NLTK Library for sentiment analysis. However, upon inspecting the results, the authors were dissatisfied with the automated classification outcomes. Consequently, they opted for manual qualitative analysis, which enabled a more in-depth and nuanced sentiment exploration (Haque et al., 2022; Rajapakse et al., 2022; Strauss & Corbin, 1997). Additionally, news articles were used because news media in democracies are considered to accurately report on public opinion (Beckers, 2020), providing insight into the audiences' topics of debate.

Consequently, the news portals included in the analysis were carefully selected to ensure that all dominant media agencies in the Czech Republic (namely MAFRA, The Czech News Center, Economia, public sphere, and independent) were covered and reflected fairly. The selection process respected the most variance of mass media audiences to be included. Based on the complex selection process, eight relevant news media were selected for the analysis: **Novinky.cz** (MAFRA), **Seznam Zprávy** (Seznam.cz), **Hospodářské noviny** (Economia), **Blesk.cz** (The Czech News Center), **Deník N** (independent), **Forbes** (owned by a Slovak agency MediaRey SE under the U.S. Forbes trademark), **ČT 24** (public), ***Chip***[83] (Burda International). The selection process also included criteria such as publishing frequency (daily, weekly, monthly), commercial or public spheres, serious press or tabloids, and newspapers or magazines. Lastly, the readership of news portals was considered to include the most popular daily press newspaper (Blesk.cz) and the most popular news internet platform (Seznam Zprávy), as reported by MediaGuru in 2021 (Vojtěchovská, 2022). Therefore, the selection should create balanced reporting regarding funding, sponsorships, political views, and target audience. Articles were collected for three months (February - April 2023). All articles that included the keyword 'ChatGPT' were content analyzed (*N* = 201) and are listed in **Appendix G**.

---

[83] Czech thematic magazine focused on information technology.

Table 6

Codes of the News Articles for Content Analysis

Code	Definition *'This code was used when the article talks…'*	Absolute Occurrence
Education	… about the impact of ChatGPT on education / academia	26
Healthcare	… about the impact of ChatGPT on healthcare	9
Fear	… about ChatGPT pictured as an AGI, humanoid, a cause of losing humanity and dystopia	38
Job Market	… about the impact of ChatGPT on the job market	47
Disinformation	… about ChatGPT in relation to disinformation	20
Competition	... about the GenAI competition among Tech companies	70
Regulation	… about regulation of ChatGPT	34

Note. The total number of codes occurrence was *N* = 244. Articles could have been assigned one, zero, or more codes (up to 8). A total of 201 Czech news articles from February to April 2023, including the keyword 'ChatGPT', were analyzed via content analysis. The complete matrix of all articles with their assigned codes can be found in the Supplementary Materials.

4.3.1. Content Analysis Results

COMPETITION. Reporting on the competitive race among BigTech companies in releasing the most powerful chatbots appeared in news articles the most often (in almost 35% of the analyzed articles). The articles usually reported on the dynamic competitive 'arms race' and the assumed business moves. Some articles also only informed about the launch of the new chatbot, such as Sberbanks' GigaChat (Russia), Alphabet's (Google's) Bard (U.S.), Elon Musk's TruthGPT (U.S.), Microsoft's Bing (U.S.), Baidu's ERNIE (China), and Amazon's Bedrock (U.S.).

> "In December, Google declared an internal alert ("code red") triggered by the public availability of the rival ChatGPT, and almost simultaneously announced the layoff of six percent of its employees (about 12,000 people). Similarly, Microsoft is laying off 10,000 people or five percent of employees. Both companies are obviously cutting back on redundant activities and betting as much as possible on AI development." [70]

The availability of the released chatbots in the Czech Republic and in the Czech language was also reflected in the media:

> “Google is much more conservative than Microsoft and offers its Bard system only in English, while Bing AI cheerfully speaks Czech. Because of this, Google restricts access to its system only to users from the United States and Great Britain.” [176]

JOB MARKET. The second most occurring topic (in 23.4% of analyzed articles) was the labor market transformation due to advanced GenAI models. Most of the articles talked about concerns about certain job positions, such as translators, journalists, radio moderators, scientists, programmers, lawyers (and the first legal AI chatbot Harvey[84]), illustrators, writers, and the fields of marketing, customer care, and grant agencies. The situation was pictured as inevitable based on two repeatedly mentioned market labor analyses: one from OpenAI[85] and the other from Goldman Sachs,[86] warning from 300 million full-time jobs dismissal due to automation.

> “The point is that AI doesn't have to completely replace the reception desk, but the receptionist should stop being a person who just picks out the ID card, writes it down and issues a room card. Any robot can do this without any problems. … This is also the case with creative jobs. If you work on a computer more than 80% of the time, and do not add any added value to it, just fill in spreadsheets, then AI will replace you with 100%.” [19]

Although inevitable, the job market situation was presented with concerns and potential. The first was represented by the need for workers’ requalification, especially for the ‘average-performance’ workers who cannot adapt to the GenAI generating higher-level content of their work. Another perspective was that this cheap content creation would lead to lower wages and salaries rather than workers’ direct dismissal. On a positive note, the articles talked about new positions being created (especially where the increase of effectivity is required), more exciting content of work with the ‘boring’ tasks being automated, shortened office hours, and essentially GenAI as a ‘helper or assistant’ rather than ‘substitution’ (such as Microsoft renewed Clippy powered by GPT-4).[87] Generally, the development of the labor market was presented as threatening yet inconclusive in its impacts.

---

[84] https://www.harvey.ai/

[85] Elondou, T., Manning, S., Mishkin, P., & Rock, D. (2023). *GPTs are GPTs: An early look at the labor market impact potential of large language models*. ArXiv:2303.10130. https://doi.org/10.48550/arXiv.2303.10130

[86] Hatzius, J., Briggs, J., Kodnani, D. & Pierdomenico, G. (2023, March 23). *The potentially large effects of artificial intelligence on economic growth (Briggs/Kodnani).* Goldman Sachs. https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf

[87] https://appsource.microsoft.com/en-us/product/web-apps/wetransactio1669903183521.clippybywetransact?tab=overview&exp=kyyw

Another dominant topic in the articles was the unconvincing quality of ChatGPT’s generated content:

> “He [ChatGPT] is like a new junior colleague with minimal experience but huge potential. You want him on the team, but at the same time you are afraid that he will outgrow you. "His work is formally faultless, he just 'lies' completely unabashedly - he mixes truth and fiction better than many political marketers, and when you catch him, he humbly apologizes to you, and then lies again. You have no leverage over him, he doesn't care about you at all.” [5]

> “At the beginning it won't even do better than us. Newspaper articles written by AI will be a little boring, theater plays rather clichéd, thriller plots unoriginal, official and business correspondence unnecessarily long-winded, and so on. This trend has been valid since the beginning of the industrial revolution, it is not specific to AI or digital technologies: technical progress usually brings slightly lower quality at a significantly lower cost.” [75]

FEAR. The fear of advanced AI systems is presented paradoxically in the Czech media, appearing in 18.9% of analyzed articles. There are two contra-intuitive perspectives presented. The first one criticizes the evoked fear related to ChatGPT and GenAI as purposeful to make people worry and expect the irreversible and uncontrolled development of AI systems. This perspective criticizes the calls for the pause of LLM development for the choice of language, which exaggerates the potential of GenAI and artificially creates the narrative of destroying humanity.

> “At the root this is the fear that we humans too will be transformed into something machine-like, into one particle just to keep the whole system running. And that something that we usually associate especially with the themes of personal freedom or creativity will be lost.” [123]

While some articles criticize it, others add to the exact fearful narrative building. Some intentionally via clickbait headlines and content about ‘uncontrollable AI’ that aims to destroy all humans (e.g., the ‘Furby’ story [151, 188]), anthropomorphizing it (e.g., “we are the guinea pigs of artificial intelligence” [119]), or complementing the articles with ‘AI humanoid’ pictures (see selected example in **Figure 9**). Other fearful narrative is also created via reporting on ChatGPT’s generative abilities, calls for a pause in development, need for regulation, and job dismissal without providing sufficient context, such as ChatGPT’s statistical nature. Only three articles out of the 38 coded to the category ‘Fear’ contain a sufficient technological explanation of ChatGPT’s architecture, providing the readers with enough information to gain a meaningful understanding (in total, nine articles out of 201 include a sufficient tech explanation).

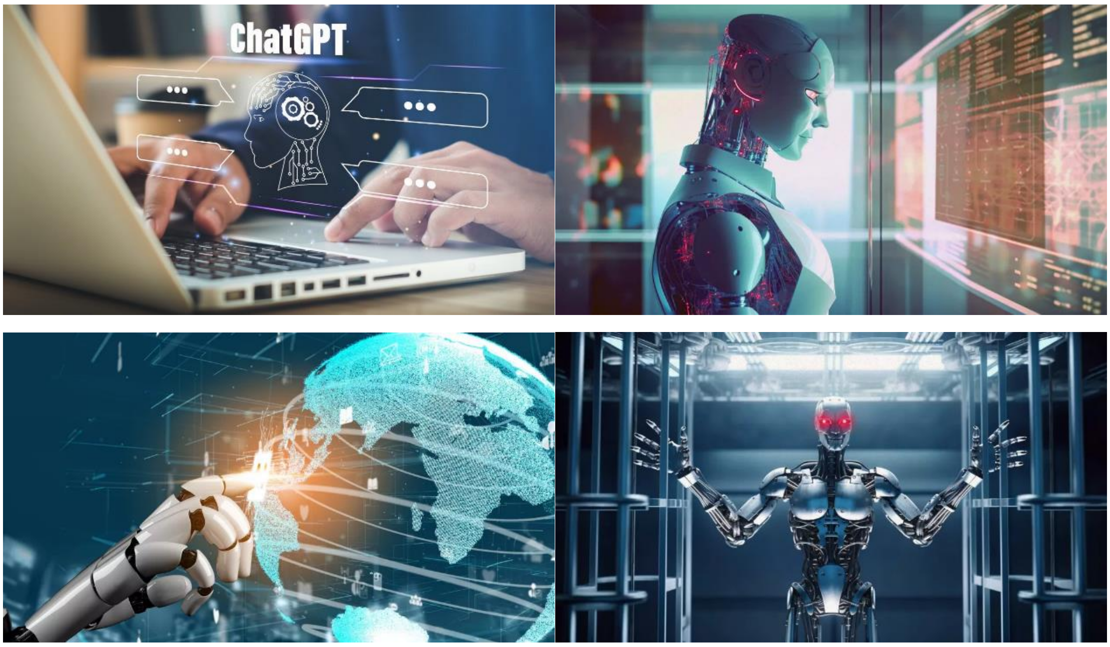


Figure 9. Illustrative images from four selected Czech news articles containing the keyword ‘ChatGPT’, collected from February to April 2023. The images were selected purposefully to demonstrate the AI humanoid portrait painted in the Czech media. Article sources: [171, 128, 126, 129].

REGULATION. Three regulatory topics resonated in the Czech media. Firstly, it was the call for the six months pause in AI development, to which the public opinion is inconclusive. While some perceive it as commercial opportunism and exaggerating the problem, others perceive it as an opportunity for the regulatory bodies to catch up and minimize the current developmental asymmetry between Big Tech companies and regulators. The second topic is the national bans of ChatGPT, such as in China or Italy, and the Czech ‘waiting out’ strategy to see what other countries and regulatory bodies will do. Finally, the role of the EU in regulating GenAI was often discussed. It is viewed as ‘AI vs European regulations,’ where the EU is supposed to play a central regulatory role in protecting users from their behavior data collection and assessment, considering the GDPR and AI Act (in the making since 2021). While not everyone has faith in the impact of the EU regulations, the whole Czech Republic relies on it and awaits its incoming.

> “Artificial intelligence will shape our everyday life, our culture and science with its creative results. Europe will probably have to play the role of a bystander, whose timid attempts at regulation in the form of a regulation on artificial intelligence will fade away into bureaucratic nirvana.” [30]

> “The MEPs[88] will probably propose an artificial intelligence regulation draft this week. This would then be followed by negotiations on the final form with the member states. All aim to approve it by the beginning of 2024. "We are dealing with a pioneering technology and using it means introducing clear rules as soon as possible," the Financial Times quoted one of the rapporteurs, MEP Dragoș Tudorache.” [193]

> “I trust the regulators to do something about it. If they said: Dear big technology companies, you have two years and then tracking will be illegal! They would certainly be able to handle it.” [158]

EDUCATION. In education, ChatGPT is perceived as a positive challenge, opportunity, and duty towards the students to prepare them for the transformation of the labor and skills market. ChatGPT in education and academia is not seen as a threat but as the area with the most significant opportunity to take advantage of ChatGPT. Supposedly the open-minded and positive attitude towards ChatGPT in education is due to the absence of teachers’ job dismissal fear and the general lack of personnel in education. LLMs are expected to lower the burden on teachers via personalization, scalability, and adaptability to the fast-changing environment.

> “My field is education and pedagogy. A journalist asked me whether AI is a threat or a challenge for schools. It is a threat only to those who perceive it as a threat. It's definitely a challenge, bigger than covid was. But above all, it is an obligation – an obligation to teach pupils and students how to use AI effectively.” [179]

The school representatives do not fear ChatGPT nor the ‘aigiarism’ (plagiarism using GenAI):

> “In the Czech education system, according to Zajíček, the practice of testing students prevents the abuse of chatbots. In contrast to, for example, Anglo-American education, writing essays is not a common practice in Czech schools, pupils and students in the Czech Republic usually get their grades from oral examinations or tests written on paper.” [38]

Therefore, there is no intention to ban the usage of ChatGPT in the educational sector. Conversely, discussions among experts and school representatives are ongoing and dedicated working groups are being formed to develop an official standpoint on involving the chatbots in the system. The greatest challenge of the Czech educational system is to rapidly adapt to reflect the new era and start teaching technological awareness, AI tools, and critical thinking at the primary and secondary

---

[88] A member of the European Parliament (MEP).

levels of education. The hype that ChatGPT has evoked should be transformed into AI literacy education. Some high schools and universities have already press released their progressive statements allowing the usage of ChatGPT while maintaining the standard plagiarism policies. The ban on the chatbots' use or any re-formulation of how the students are currently being assessed is not on the table. However, this open-minded approach towards including the recent AI technologies in education is not a nationwide bylaw but a bottom-up initiative.

> "Teaching artificial intelligence and how to work effectively with it will also be a challenge for Czech education. Zeman mentions the AI literacy. It will involve knowledge of entering correct commands to artificial intelligence, knowledge of its functions and awareness of technological limits. In the case of insufficient education in this area … there is a risk of the creation of a new kind of knowledge gap, which will divide students into 'prospective' and 'disadvantaged' ones." [43]

> "… the need to focus more on the technological innovations in education, so that education reflects the times we live in … To teach the young generation how to safely use the new technological tools like ChatGPT and applied critical thinking. But that's what agree to here in our filter bubble. And then we come across the reality of contemporary Czech education. I don't see the willingness on the part of the schools..." [11]

DISINFORMATION. The topic of disinformation (or misinformation) concerning ChatGPT only appeared in 20 articles (9.9%). Typically, the articles reported on the 'scandals' that made the international news, such as the fabricated interview with Michael Schumacher in the German daily news [65] or the Australian mayor's lawsuit against OpenAI after ChatGPT generated false information about him [56, 94]. However, only a few articles discussed the security aspect and the potential for human manipulation that the AI text generation increases.

> "Unfortunately, people are very easily influenced. Take, for example, chain e-mails, a specialty of Central Europe. This is a kind of DDoS attack on human brains (DDoS is a type of cyberattack that aims to make the Internet service unavailable to ordinary users by flooding it - editor's note). … They serve them to people through text, which can easily be created using language models thanks to artificial intelligence. Today it is already becoming difficult to recognize what is created by a real person and what is machine." [158]

HEALTHCARE. Although the educational sector was widely represented in the news, healthcare was only coded in nine out of 201 articles. The few articles recognized the importance of AI for future healthcare development and its potential for automation, diagnostics, telemedicine, drug development, and medical education. However, Czech healthcare is reportedly conservative and inflexible, with no prospect for innovation.

> “There are two areas where AI will save the planet. The first is healthcare, where there will be a radical reduction in the healthcare cost for hundreds of millions of people who cannot access it today. Today, doctors still look at records and results, over and over and over again, unnecessarily. There is huge potential for automation and diagnostics. If you know the diagnosis in advance, the cost of treatment will be just fractional.” [158]

4.4. Discussion

The hypothesis that Czechs will be close-minded, cautious, and techno-sceptic towards ChatGPT has been only partially confirmed. While Czech people indeed have a low adoption rate of ChatGPT, it is not an inner scepticism that prevents them from adopting the technology. Exactly the opposite, Czech people are traditionally fond of new technological innovations and present a warm testing environment for deploying new technologies, as demonstrated, for instance, by the extensive field testing in Prague conducted by Samsung (2012). However, what prevents Czech people from integrating ChatGPT is the artificially evoked fear of job loss and human destruction. While unharmful tools, such as phones and LTE networks are widely adopted, ChatGPT is leveraged as a *product* rather than a *tool*.

The populistic 'fear shouting' is assumed to have three purposes: (a) to leverage the general worldwide interest to market the technology (e.g., as part of a sci-fi fantasy or via clickbait), (b) to submerge the more significant issues related to LLMs, and (c) shift the responsibility for the models' outcomes away from the developing companies by creating the 'apocalyptic' conscious ChatGPT narrative. In other words, assigning ChatGPT's human-like traits make it 'responsible' for its own actions. Even the former White House Tech adviser and computational philosopher Suresh Venkatasubramanian[89] labeled the exaggerated threatening claims as an "organized campaign of fearmongering" (Goldman, 2023b). The populistic-evoked fear is not unheard of to the Czech people; quite the opposite. Sociologist David Klimeš identifies the prevailing populist-induced fear as one of the key problems of the stagnating Czech country (Klimeš, 2020). According to Klimeš, the populist-induced fear reappears with every Czech election, where the populistic fear meets with a realistic idealism in a cultural fight. The Czech media rapidly adapted to the Czech situation and developed a 'populistic algorithm,' preferring conflicting opinions, emotions, and radical and bold claims over consensus, facts, and rational and professional debates (Klimeš, 2020). This was reflected in the most represented topics in the Czech news articles relating to ChatGPT: 'arms race' competition, job dismissal, and populistic fear with images of apocalypse and humanoids. Notably, these images were not present in all media types. Nevertheless, most of them appeared in the Seznam Zprávy news portal, the most popular news internet platform among the Czech population (Vojtěchovská, 2022). Given the Czech sensitivity to artificially evoked fear, their generally low AI literacy, and the absence of a regulatory body, it is not surprising that ChatGPT is dominantly unadopted. While only 5% of people interacted with ChatGPT, six out of ten support the call for a pause in development (Ipsos, 2023).

---

[89] https://www.linkedin.com/in/suresh-venkatasubramanian-233b751/

However, it is not due to the 'Scepticism' towards innovation but rather the half-truths fed to the public debates and the hope for European authorities to take control. Consequently, the Czechs are identified by the 'Romanticism' category (**Figure 7**; Kerschner & Ehlers, 2016), represented by its ambiguity towards technologies (inconclusive opinion and confusion) and unclear governing tendencies. The conceptual ambiguity trend is in accordance with other findings about Czech technological attitudes (Blažek & Uhlíř, 2007; Kershner & Ehlers, 2016).

While the topic of fear is widely represented in the Czech public sphere, the absence of certain legal issues does not match the expectations. Topics such as disinformation, privacy, and misuse of personal data concerns could be expected, given the European regulatory nature. While the experts are concerned about these legal and regulatory problems, which supports the second hypothesis of them belonging to the AI Safety ideology (Tiku, 2023), it does not propagate to the general population discourse. Czech people show a lack of concern about these topics (e.g., the topic of disinformation in only 9.9% of analyzed articles) is in accordance with previous studies related to concerns about autonomous vehicles (Gabrhel et al., 2019; Zámečník et al., 2021).

On the other hand, the topic of regulation is more prevalent, appearing in almost 17% of articles and dominantly in interviews with experts. Nevertheless, the willingness to attend to the ChatGPT regulations is low, and the responsibility is shifted to the EU level. The Czech Prime Minister Petr Fiala has just signed an open letter (issued by six other Prime Ministers) addressed to Big Tech companies. This open letter urgently asks Big Tech companies to take responsibility and control over disinformation being shared on social media platforms (Government of the Czech Republic, 2023). After years of not dealing with this issue, this letter was issued on March 29, 2023. This demonstrates the accountability battle between policymakers and systems developers, shifting responsibility from one to the other. Additionally, it shows how stagnating and protracted the official response to technological trends is in the Czech Republic, which was reflected in the analyzed data. Therefore, the current Czech attempts to implement ChatGPT into business are uncontrolled. Since the whole EU is lagging on the LLMs development, the Czech people play the role of passive bystanders following the competitive developmental race between Big Tech companies. The interest in the topic is not surprising, given the at-risk dominance of Google on the search engine market after almost ten years (Google have shared more than 80% of the market since 2015; Bianchi, 2023b). While the topic of Big Tech competition is prevalent, the concerns about their business cases and monetization practice are missing in the public debates.

The last dominant topic in both analyses was education, as opposed to healthcare, which almost did not make the news. Supposedly, it is due to the Czech healthcare quality being at the top European level, while education lags in many aspects (Klimeš, 2020). The topic of healthcare was not represented in the public debates. Supposedly, it was because AI innovations are not a new technology for the health system. Therefore, ChatGPT did not create hype there.[90] Successful AI systems are already implemented, such as the Czech diagnostic radiology system Carebot.[91] However, education deals with fundamental problems, such as educational inequalities, large differences between regions, severely underpaid teachers, and a lack of qualified personnel (Klimeš, 2020). Due to these issues, educational innovation is almost non-existent, and ChatGPT represents an opportunity to improve of the stagnating system. Additionally, representatives in education had to react to the students' rapid adoption of ChatGPT (the highest significant adoption rates are among students and people under 24 years old; Ipsos, 2023; NMS Market Search, 2023). In response, the educational representatives, including teachers and researchers from universities, introduced bottom-up working groups, initiatives, and statements reflecting the currently evoked situation.

Two aspects were essential for such a progressive position that many educational institutions have taken. Firstly, it is a large degree of autonomy, thanks to which most decisions were historically in the control of individual schools. This autonomy is only being restricted in the last years, with the integration of nationwide systemized high schools' entering exams (since 2017) and final leaving exams called 'Maturita' (since 2011). The second key aspect is the precedent set up by the COVID-19 pandemic, where the schools had to react rapidly to the fast-changing situation with no official government statement to guide them at the crucial beginning (Klimeš, 2020). Given these prepositions, the Czech field of education has adopted a positive and open-minded attitude towards using ChatGPT to prepare future generations for the rapid transformation of work-relevant competencies. The Czech educational sector is the only one that arguably successfully adapted to the ChatGPT disruption.

---

[90] Notably, there are still issues to the Czech healthcare system, which were shown especially during the COVID-19 pandemic (Klimeš, 2020). However, the quality the care is high in the Czech Republic.

[91] https://www.carebot.com/

5. The Final Chapter: Conclusion

This thesis leveraged knowledge from various scientific disciplines to explore the role and broader behavior of ChatGPT in society with an extension to the Czech context. It theoretically explored the impact of human decisions on the model architecture and capabilities and the technology's impact on society via systematically analyzing the risks and potential. While on the architectural level, the model is nothing unseen, it created a disruptive worldwide response. Therefore, while there is potential, many concerns are disturbingly unattended and underrepresented in public debates. The specifics change with the different social environments and the attitude towards technology. Therefore, this thesis also qualitatively analyzed the Czech topics of public debates about ChatGPT via thematic analysis of four experts' interviews and content analysis of 201 news articles. Generally, half-truths, lack of context, and populistically evoked fear are presented instead of AI literacy education. The public debates concentrate on the Big Tech competition, job dismissal, and unprecedented 'consciousness' of the system instead of questioning the responsibility for the system, its ecological burden, the disinformation issue, its business case, or monetization practice. It evokes feelings of purposeful attention-shifting and clever marketing since the Big Tech companies have adopted the mindset of 'ask for forgiveness, not permission,' optimizing for progress over harm mitigation.

Nevertheless, GenAI technologies are the next big thing. The responsibility for mindfully implementing them in society and adapting them to societal needs should be addressed. However, currently, neither the developing companies nor the elected politicians and authorities are admitting their responsibility share. While this varies in different social environments, the Czech authorities are certainly not addressing the issue. The only Czech sector that 'keeps a finger on the pulse of the time,' which means to keep pace with and be aware of the latest progress, is education. While education in the Czech Republic is not *keeping* pace, it at least figuratively *entered the race*. It aims to provide Generation Z (GenZ, the *internet* generation) with flexible learning and collaborative skills, which will be extremely valuable in the ever- and fast-changing environment. Given the GenAI-evoked transformation of society that will come, such as with the job market and relevant work skills, one can imagine that the future generation should be called *Gen*AI. It is no longer the internet generation but rather the generation of *(generative) artificial intelligence*. Therefore, the need for adaptability and inner stress resilience may be even greater than Yuval Noah Harari (2018) argued in his influential book 21 Lessons for the 21st Century. Consequently, leveraging the recent hype around AI systems should focus on educating laypeople and equipping them with abilities instead of spooking them with the ideas of half-truths and apocalypse.

This thesis is a snapshot in time. It was initiated in February 2023 and completed by the end of May 2023. Given that in this short period, GPT-4 was released, ChatGPT was banned in Italy, scandals with data leakage occurred, and more, it only reflects the current situation of ChatGPT's risks, benefits, impacts, and transformation. Due to this dynamic situational development, pieces of information in this thesis or even parts of the work may be outdated soon. However, the thesis aimed to present a fresh perspective into the societal impact analysis with the: (a) outmost recent LLMs development, (b) their impact on society, and (c) Czech cultural aspects and context.

Consequently, the thesis is unique in its holistic approach to exploring the phenomena of ChatGPT from many perspectives. To the author's knowledge, this thesis is the first to analyze ChatGPT in such an interdisciplinary and complex approach. Providing the complex context, including the model's evolution and architecture, is the first necessary step to a more profound understanding of its societal impact. Analyzing the societal influence and impact is the next step required to safely mitigate the risks and mindfully adopt the new technology into the daily lives of the general population. Therefore, the rational acquisition of the present thesis is significantly valid and relevant in today's world, enriching the future regulatory or implementational attempts of a valuable fundamental analytical overview.

5.1. Limitations

This work is the first to explore ChatGPT within the Czech context and the Czech attitudes towards technology. Additionally, it is also novel in its extent, covering all major aspects and events related to ChatGPT since its launch. Given the broad scope of this thesis, it is not free of limitations. Firstly, little academic literature is available as the thesis was created less than half a year after ChatGPT-3.5 was launched. Given the inefficient editorial process,[92] it is too early for many peer-reviewed journals to publish. Therefore, the present theoretical analysis included non-peer-reviewed studies, and the findings should be considered accordingly. The one existing meta-analysis of ChatGPT after 2.5 months also included more than 150 papers, primarily from ArXiv[93] and SemanticScholar[94] (Leiter, 2023).

Another limitation is the uncertain extent to which the topics in news media represent public opinion. Some journalistic studies state that media in democracies often report on public opinion to which they are closely intertwined, keeping the audience informed about the dominant discourse in their population (Beckers, 2020; Deuze, 2005). However, Kalogeropoulos (2019)

---

[92] In the fields of computer sciences, social sciences, and humanities, the average editorial process takes up to 22-23 weeks (Huisman & Smits, 2017).

[93] https://arxiv.org/

[94] https://www.semanticscholar.org/

pointed out that people under the age of 35 do not use traditional news media to the extent that older generations do. Therefore, the public discourse presented in the mass media may underrepresent certain social groups. Since the methodology choice made it impossible to assess any socio-demographics, the representativeness of the data should not be assumed. Consequently, while the findings are representative of the dominant topic coverage in the Czech public sphere, they should not be presented as public discourse without providing the essential context.

Regarding the conducted analyses, the author of this thesis acknowledges their Czech origin. Given the pre-understanding, cultural bias, and personal opinions towards the Czech culture, measures were taken to maintain a reflective understanding of the author's awareness and objective findings. However, the pre-understanding may have been leveraged into a deeper data understanding, which aligns with the Erlingsson and Brysiewicz (2017) guide to content analysis. Any future work may want to consider analyzing the adoption of such disruptive technology in other social settings and compare the results to the Czech context. Additionally, future work should consider multi-modal models since it can be proposed that they will substitute for text-only ChatGPT in its success.

The last point to address is the authorship of the pieces of literature already available. Some of the articles cited throughout the thesis claimed the usage of a language model to create (a part of) the text, e.g., "The human author of this article would like to state that this entire article was written by ChatGPT" (Biswas, 2023, p.2). While it is hard to navigate, to the author of this thesis knowledge, all cited parts of the thesis were authored by a human(s), and so the thesis contains 100% pure 'human semantics.' It is WBH (Written By Human).

References

Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. G. (2020). Roles of computing in social change. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency,* 252–260. https://doi.org/10.1145/3351095.3372871

Altman, S. (2023c, February 24). *Planning for AGI and beyond.* OpenAI. https://openai.com/blog/planning-for-agi-and-beyond

Altman, S. [@sama] (2022, December 5). *ChatGPT launched on wednesday. today it crossed 1 million users!* [Tweet]. Twitter. https://twitter.com/sama/status/1599668808285028353?lang=en

Altman, S. [@sama]. (2020, July 19). *The GPT-3 hype is way too much. It's impressive (thanks for the nice compliments!) but it still has serious weaknesses* [Tweet]. Twitter. https://twitter.com/sama/status/1284922296348454913

Altman, S. [@sama]. (2023a, March 22). *unfortunately, users will not be able to access their chat history from monday 1 am PDT until monday 10 am* [Tweet]. Twitter. https://twitter.com/sama/status/1638635718691135488

Altman, S. [@sama]. (2023b, March 22). *we had a significant issue in ChatGPT due to a bug in an open source library, for which a* [Tweet]. Twitter. https://twitter.com/sama/status/1638635717462200320

Asimov, I. (1942). *Runaround. Astounding Science Fiction.* Street & Smith Publications, Inc.

Asimov, I. (1950). *I, Robot*. Gnome Press.

AtlasVPN (2023, April 5). *VPN downloads in Italy skyrocket by 400% after ChatGPT block*. AtlasVPN Blog. https://atlasvpn.com/blog/vpn-downloads-in-italy-skyrocket-by-400-after-chatgpt-block

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelly, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, 10.1001/jamainternmed.2023.1838

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations.* https://arxiv.org/pdf/1409.0473.pdf

Baidoo-Anu, D., & Ansah, L. O. (2023). *Educating in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning*. SSRN. http://dx.doi.org/10.2139/ssrn.4337484

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1):37–46.

Barwise, J. (1977). An introduction to first-order logic. In J. Barwise (Ed.), *Handbook of Mathematical Logic* (pp. 5-46). North Holland Publishing Company.

Baschez, N. [@nbashaw] (2022, December 6). *ChatGPT probably cost $1–3 million to run (in its first five days of existence) We know they have >1m* [Image attached] [Tweet]. Twitter. https://twitter.com/nbashaw/status/1599949577821786113?s=46&t=NZTpGNbA7kBNDNGLbpk3bw

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report, 13*(4), 544-559. https://doi.org/10.46743/2160-3715/2008.1573

BBVA Foundation (2012). BBVA Foundation International Study on Scientific Culture: Understadning Science. BBVA Foundation. https://www.fbbva.es/wp-content/uploads/2017/05/dat/Understandingsciencenotalarga.pdf

Beckers, K. (2020). The voice of the people in the news: A content analysis of public opinion displays in routine and election news. *Journalism Studies, 21*(15), 2078-2095. https://doi.org/10.1080/1461670X.2020.1809498

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. 10.18653/v1/2020.acl-main.463

Bender, E. M., Gebru, T., McMillan-Major, A., Schmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,* 610-623. 10.1145/3442188.3445922

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research, 3*, 1137–1155.

Bezos, Jeff (2006, September 27). *Opening Keynote and Keynote Interview* [Conference presentation]. MIT World -Special Events and Lectures. http://techtv.mit.edu/videos/16180-opening-keynote-and-keynote-interview-with-jeff-bezos

Bianchi, T. (2023a, February 22). *Share of adults in the United States who have seen, read, or heard about ChatGPT as of February 2023, by generation*. Statista. https://www.statista.com/statistics/1367851/us-adults-awareness-of-chatgpt-by-generation/

Bianchi, T. (2023b, May 10). *Global market share of leading desktop search engines 2015-2023.* Statista. https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

Biswas, S. (2023). ChatGPT and the future of medical writing. *Radiology, 307*(2). https://doi.org/10.1148/radiol.223312

Blake, A. (2023, April 4). *Someone just used ChatGPT to generate free Windows keys.* Digital Trends. https://www.digitaltrends.com/computing/chatgpt-generates-free-windows-keys/

Blažek, J., & Uhlíř, D. (2007). Regional innovation policies in the Czech Republic and the case of Prague: An emerging role of a regional level? *European Planning Studies, 15*(7), 871-888. https://doi.org/10.1080/09654310701356175

Bolton, E., Hall, D., Yasunaga, M., Lee, T., Manning, C., & Liang, P. (2022). *Stanford CRFM introduces PubMedGPT 2.7B.* Stanford University. https://hai.stanford.edu/news/stanford-crfm-introduces-pubmedgpt-27b

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2021). *On the opportunities and risks of Foundation models*. ArXiv:2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Borji, A. (2023). *A categorical archive of ChatGPT failures*. ArXiv:2302.03494. https://arxiv.org/abs/2302.03494

Brandl, R., & Ellis, C. (2023, March 31). *Survey: ChatGPT and AI content. Can people tell the difference?* Tooltester. https://www.tooltester.com/en/blog/chatgpt-survey-can-people-tell-the-difference/#Methodology

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. *Interactions, 25*(5), 38-43. 10.1145/3236669

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. http://dx.doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. Sage.

Breazeal, C., & Scassellati, B. (2000). Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior, 8*(1):49–74. https://doi.org/10.1177/105971230000800104

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482-1493. 10.1037//0278-7393.22.6.1482

Brockman, G., Sutskever, I., & OpenAI (2019, March 11). *OpenaAI LP.* OpenAI. https://openai.com/blog/openai-lp#GregBrockman

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyan, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., …, & Amodei, D. (2020). *Language models are few-shot learners*. ArXiv: 2005.14165. https://doi.org/10.48550/arXiv.2005.14165

Browne, R. (2023, January 10). *Microsoft reportedly plans to invest $10 billion in creator of buzzy A.I. tool ChatGPT.* CNBC. https://www.cnbc.com/2023/01/10/microsoft-to-invest-10-billion-in-chatgpt-creator-openai-report-says.html

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E. Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of artificial general intelligence: Early experiments with GPT-4*. ArXiv:2303.12712. https://doi.org/10.48550/arXiv.2303.12712

Buchanan, B., Lohn, A., Musser, M., & Katerina, S. (2021). *Truth, lies, and truth, lied and automation: How language models could change disinformation automation*. Center for Security and Emerging Technology. https://cset.georgetown.edu/publication/truth-lies-and-automation/

Buchholz, K. (2023, January 24). *ChatGPT Sprints to One Million Users*. Statista. https://www.statista.com/chart/29174/time-to-one-million-users/

Burgess, M. (2023, April 13). *The hacking of ChatGPT is just getting started*. WIRED. https://www.wired.com/story/chatgpt-jailbreak-generative-ai-hacking/

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine, 9*(2), 48-57. 10.1109/MCI.2014.2307227

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, U., Oprea, A., & Raffel, C. (2021). *Extracting training data from large language models*. ArXiv:2012.07805. https://doi.org/10.48550/arXiv.2012.07805

Čapek, K. (1921). R.U.R. [Rossum's Universal Robots].

Carmon, Y., Raghunathan, A., Schmidt, L., Duchi, J. C., & Liang, P. S. (2019). Unlabeled data improves adversarial robustness. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019).* https://papers.nips.cc/paper_files/paper/2019/hash/32e0bd1497aa43e02a42f47d9d6515ad-Abstract.html

Casilli, A. A. (2021). Waiting for robots: The ever-elusive myth of automation and the global exploitation of digital labor. *Sociologias, 23*(5), 112-133. 10.1590/15174522-114092

Casilli, A. A. (2022, September 19-23). *Digital Labor, Platforms, and AI* [Conference presentation]. Digital Humanism Summer School, Technical University, Vienna, Austria. https://owncloud.tuwien.ac.at/index.php/s/TyuHTaZtLVI2hyh

Casilli, A. A., & Posada, J. (2019). The platformization of labor and society. In M. Graham & W. H. Dutton (Eds), *Society and the Internet. How Networks of Information and Communication are Changing Our Lives* (2nd ed., pp.293-306). Oxford University Press.

Ceccato, S. (1967). Correlational analysis and mechanical translation. In A. D. Booth (Ed.), *Machine Translation* (pp. 77-135). North Holland Publishing Company.

CESIE (2022, January 31). *What is the true cost of disinformation*? CESIE. https://cesie.org/en/youth/true-cost-of-disinformation-mega/

Cheng, J., Dong, L., & Lapata, M. (2016). *Long short-term memory-networks for machine reading*. ArXiv:1601.06733. https://doi.org/10.48550/arXiv.1601.06733

Chomsky, N. (1957). *Syntactic structures*. De Gruyter Mouton. https://doi.org/10.1515/9783112316009

Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. In *Proceedings of Advances in Neural Information Processing Systems 30 (NIPS 2017).* https://papers.nips.cc/paper_files/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf

Chuan, Li (2020, June 3). *OpenAI's GPT-3 Language Model: A Technical Overview*. Lambda. https://lambdalabs.com/blog/demystifying-gpt-3

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204. https://doi.org/10.1017/S0140525X12000477

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

Collobert, R., & Weston., J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, 160–167. https://doi.org/10.1145/1390156.1390177

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology*. University of Chicago Press.

Cote, K. A., Etienee, L., & Campbell, K. B. (2001). Neurophysiological evidence for the detection of external stimuli during sleep. *Sleep, 24*(7), 791-803. https://pubmed.ncbi.nlm.nih.gov/11683482/

Crick, F., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex, 8*(2), 97-107. 10.1093/cercor/8.2.97

D'Amico, R. S., White, T., Shah, H., & Langer, D. (2023). I asked a ChatGPT to write an editorial about how we can incorporate chatbots into neurosurgical research and patient care…*General Neurosurgery, 92*(4), 663-664. 10.1227/neu.0000000000002414

Dai, A. M., & Le, Q. V. (2015). *Semi-supervised sequence learning*. ArXiv. https://doi.org/10.48550/arXiv.1511.01432

Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering, 27*(1), 113–118. https://doi.org/10.1017/S1351324920000601

Darcy, A. (2023, March 1). *Why generative AI is not yet ready for mental healthcare* [Article]. LinkedIn. https://www.linkedin.com/pulse/why-generative-ai-yet-ready-mental-healthcare-alison-darcy/?trackingId=KsWtz%2B49lX8Ytoo8IyVNTQ%3D%3D

Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of human-level bonds established with a digital conversational agent: Cross-sectional, retrospective observational study. *Journal of Medical Internet Research Formative Research, 5*(5), e27868. 10.2196/27868.

Davidson, H. (2023, February 23). *'Political propaganda': China clamps down on access to ChatGPT*. The Guardian. https://www.theguardian.com/technology/2023/feb/23/china-chatgpt-clamp-down-propaganda

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & Society, 35*, 917-926. https://doi.org/10.1007/s00146-020-00960-w

De Vynck, G., Lerman, R., & Tiku, N. (2023, February 16). *Microsoft's AI chatbot is going off the rails.* The Washington Post. https://www.washingtonpost.com/technology/2023/02/16/microsoft-bing-ai-chatbot-sydney/

Dehaene, S., & Changeux, J-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*(2), 200-227. https://doi.org/10.1016/j.neuron.2011.03.018

Deng, J., Dong, W., Socher, R., Lu, L. J., Li, K., & Li, F. F. (2009). ImageNet: a large-scale hierarchical image database. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. 10.1109/CVPR.2009.5206848

Descartes, R. (1637). *Discourse on Method: Part One*. In Cottingham et al. (Eds.), The Philosophical Writings of Descartes. Cambridge: Cambridge University Press.

Deuze, M. (2005). What is journalism? Professional identity and ideology of journalists reconsidered. *Journalism, 6*(4), 442–464. 10.1177/1464884905056815

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. ArXiv:1810.04805. https://doi.org/10.48550/arXiv.1810.04805

Digital Humanism Initiative [DHI] (2023). *ChatGPT – a catalyst for what kind of future? Statement of the Digital Humanism Initiative, March 2023*. Retrieved from: https://owncloud.tuwien.ac.at/index.php/s/k7iz0PELf0V0vFf

Elondou, T., Manning, S., Mishkin, P., & Rock, D. (2023). *GPTs are GPTs: An early look at the labor market impact potential of large language models*. ArXiv:2303.10130. https://doi.org/10.48550/arXiv.2303.10130

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Science, 5*(1), 16-25. 10.1016/s1364-6613(00)01568-0

Erlingsson, C., & Brysiewicz, P. (2017). A hands-on guide to doing content analysis. *African Journal of Emergency Medicine, 7*(3). 93-99. https://doi.org/10.1016/j.afjem.2017.08.001

European Commission (2012). *Special Eurobarometer 382*. *Public attitudes towards robots*. European Commission. https://www.ab.gov.tr/files/ardb/evt/Public_attitudes_toward_robots_2012.pdf

European Commission (2013). *Special Eurobarometer 401*. Responsible research and innovation (RRI), science and technology. European Commission. https://ketlib.lib.unipi.gr/xmlui/bitstream/handle/ket/1102/Special%20Eurobarometer%20401.pdf?sequence=2

European Commission (2022). *The digital economy and society index*. European Commission. https://digital-strategy.ec.europa.eu/en/policies/desi

Fall, J. G. [@jacksonfall] (2023, March 15). *I gave GPT-4 a budget of $100 and told it to make as much money as possible. I'm acting as* [Image attached] [Moment]. Twitter. https://twitter.com/jacksonfall/status/1636107218859745286

Fast, E., & Horvitz, E. (2017). Long-term trends in the public perception of artificial intelligence. *31st AAAI conference on Artificial Intelligence*, *31*(1), 4–9. https://doi.org/10.1609/aaai.v31i1.10635

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*, 681-694. https://doi.org/10.1007/s11023-020-09548-1

Ford, B. J. (2023, March). *This is nonsensical. These bots have access to digitised information everywhere. Of course they pass exams! It's like a human* [Comment]. LinkedIn. https://www.linkedin.com/feed/update/urn:li:activity:7041723354612957184?commentUrn=urn%3Ali%3Acomment%3A%28activity%3A7041723354612957184%2C7042452628118609922%29

Fox, E. J. (2015, December 15). *Sam Altman on his plan t keep A.I. out of he hands of the "Bad Guys".* Vanity Fair. https://www.vanityfair.com/news/2015/12/sam-altman-elon-musk-openai

Fridman, L., & Altman, S. (Host). (2023, March 25). Sam Altman: OpenAI CEO on GPT-4, ChatGPT, and the Future of AI (Np. 367) [Audio Podcast]. In *Lex Fridman Podcast*. https://lexfridman.com/sam-altman/

Fridman, L., & Tegmark, M. (Host). (2023, April 13). Max Tegmark: The Case for halting AI development (No. 371) [Audio Podcast]. In *Lex Fridman Podcast*. https://open.spotify.com/episode/5al9TwC3RihfDqMkyqGte6

Fridman, L., & Yudkowsky, E. (Host). (2023, March 30). Eliezer Yudkowskz: dangers of AI and the end of human civilization (Np. 368) [Audio Podcast]. In *Lex Fridman Podcast*. https://open.spotify.com/episode/2g6WjOL1J1Ovm1kndXk1nt

Future of Life Institute (2023, March 22). *Pause giant AI experiment: An open letter*. Future of Life Institute. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Gabrhel, V., Ježek, S., & Havlíčková, D. (2019). Public opinion on connected and automated vehicles: the Czech context. *Transactions On Transport Sciences, 10*(2), 42-52. 10.5507/tots.2019.011

Ganguli, D., Hernandez, D., Lovitt, L., Dassarma, N., Henighan, T., Jones, A., Joseph, N., Kernion, J., Mann, B., Askell, A., Bai, Y., Chen, A., Conerly, T., Drain, D., Elhage, N., Showk, S. E., Fort, S., Hatfield-Dodds, Z., Johnson,S., …, & Clark, J. (2022). Predictability and surprise in large generative models. *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency,* 1747-1764. https://doi.org/10.1145/3531146.3533229

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2022). *Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers.* bioRxiv. https://doi.org/10.1101/2022.12.23.521610

Garante Per La Protezione Dei Dati Personali (2023, March 30). *Provvedimento del 30 marzo 2023 [9870832] [Order of 30 March 2023 [9870832]].* https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870832

Gates, B. (2023, March 21). *The age of AI has begun. Artificial intelligence is as revolutionary as mobile phones and the Internet.* GatesNotes The Blog of Bill Gates. https://www.gatesnotes.com/The-Age-of-AI-Has-Begun

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for datasets. ArXiv:1803.09010. https://doi.org/10.48550/arXiv.1803.09010

Global Silicon Valley, (2023, April 19). *The future of integrity in the brave new world of AI/GPT | ASU+GSV 2023* [Video]. YouTube. https://www.youtube.com/watch?v=TDC1ceJbdPg

Goldberg, H., Harel, M., & Malach, R. (2006). When the brain loses itself: Prefrontal inactivation during sensorimotor processing. *Neuron, 50*(2), 329-339. 10.1016/j.neuron.2006.03.015

Goldman, S. (2023a, March 15). Lightning AI CEO slams OpenAI's GPT-4 paper as 'masquerading as research'. VentureBeat. https://venturebeat.com/ai/lightning-ai-ceo-slams-openais-gpt-4-paper-as-masquerading-as-research/

Goldman, S. (2023b, March 28). *Sen. Murphy's tweets on ChatGPT spark backlash from former White House AI policy advisor*. Venture Beat. https://venturebeat.com/ai/sen-murphys-tweets-on-chatgpt-spark-backlash-from-former-white-house-ai-policy-advisor/

Goldstein, I., & Papert, S. (1977). Artificial intelligence, language, and the study of knowledge. *Cognitive Science, 1*(1), 84-123. https://doi.org/10.1207/s15516709cog0101_5

Government of the Czech Republic (2023, March 29). *Prime Minister Fiala joins the open letter of the heads of government to the heads of technology giants*. Government of the Czech Republic. https://www.vlada.cz/en/media-centrum/aktualne/prime-minister-fiala-joins-the-open-letter-of-the-heads-of-government-to-the-heads-of-technology-giants-204435/

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). *More than you've asked for: A comprehensive analysis of novel prompt injection threats to application-integrated large language models*. ArXiv:2302.12173. https://doi.org/10.48550/arXiv.2302.12173

Hansen, A. L., & Kazinnik, S. (2023). *Can ChatGPT decipher fedspeak?* Papers with Code. http://dx.doi.org/10.2139/ssrn.4399406

Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). "*I think this is the most disruptive technology" Exploring sentiments of ChatGPT early adopters using twitter data*. ArXiv:2212.05856v1. https://doi.org/10.48550/arXiv.2212.05856

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. (2018). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, *449*, 1-14. https://doi.org/10.1145/3173574.3174023

Harari, Y. N. (2018). *21 Lessons for the 21st Century* (1st ed.). Spiegel & Grau, USA.

Harnad, S. (1990). The symbol grounding problem. *Physica, 42*, 335-346.

Harrison, K., & Wang, P. (2022, December 21). *Deep dive OpenAI.* Contrary Research. https://research.contrary.com/reports/openai

Hart, B., & Risley, T. R. (1995). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4-9.

Hatzius, J., Briggs, J., Kodnani, D. & Pierdomenico, G. (2023, March 23). *The potentially large effects of artificial intelligence on economic growth (Briggs/Kodnani).* Goldman Sachs. https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf

Heaven, W. D. (2020, July 20). OpenAI's new language generator GPT-3 is shockingly good-and completely mindless. *MIT Technology Review.* https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Leaning Research, 21*(248), 1-43. https://jmlr.org/papers/v21/20-312.html

Herculano-Houzel, S. (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *PNAS, 109*, 10661-10668. https://doi.org/10.1073/pnas.1201895109

Hern, A. (2022, December 4). *AI bot ChatGPT stuns academics with essay-writing skills and usability.* The Guardian. https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hormozi, A. (2023, January 10) [@AlexHormozi]. *This is a frightening visual for me. The first dot is the amount of data Chat GPT 3 was trained* [Tweet]. Twitter. https://twitter.com/AlexHormozi/status/1612913266195587072

Huang, J., Cook, G. G., & Xie, Y. (2021). Large-scale quantitative evidence of media impact on public opinion toward China. *Humanities & Social Sciences Communication, 8*, 181. https://doi.org/10.1057/s41599-021-00846-2

Hudson, J., Orviska, M., & Hunady, J. (2019). People's attitudes to autonomous vehicles. *Transportation Research Part A: Policy and Practice, 121*, 164-176. https://doi.org/10.1016/j.tra.2018.08.018

Huisman, J., & Smits, J. (2017). Duration and quality of the peer review process: The author's perspective. *Scientometrics, 113*, 633-650. https://doi.org/10.1007/s11192-017-2310-5

Hutchinson, B., Prbhakaran, V., Denton, E., Webster, K., Zhong, Y., & Denyl, S. (2020). Social biases in NLP models as barriers for persons with disabilities. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,* 5491-5501. 10.18653/v1/2020.acl-main.487

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology, 61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Walton Family Foundation (2023, March 1). *Understanding teacher and student views on ChatGPT*. Impact Research. https://www.waltonfamilyfoundation.org/understanding-teacher-and-student-views-on-chatgpt

Ingold, T. (1995). Epilogue. Technology, language, intelligence: A reconsideration of basic concepts. In K. R. Gibson, & T. Ingold (Eds.), *Tools, Language and Cognition in Human Evolution* (pp. 449-469). Cambridge University Press.

Ipsos (2023, April 24). *Postoj české populace k umělé inteligenci je obezřetný, 6 z 10 lidí by podpořilo výzvu k pozastavení vývoje nových modelů AI* [The attitude of the Czech population towards artificial intelligence is cautious, 6 out of 10 people would support the call to pause the development of new AI models]. Ipsos. https://www.ipsos.com/cscz/postoj-ceske-populace-k-umele-inteligenci-je-obezretny-6-z-10-lidi-podporilo-vyzvu-kpozastaveni

Janssen, A., Grutzner, L., & Breitner, M. H. (2021). Why do chatbots fail? A critical success factors analysis. *42nd International Conference on Information Systems*. https://www.researchgate.net/publication/354811221_Why_do_Chatbots_fail_A_Critical_Success_Factors_Analysis

Jelinek, F., & Mercer, R. L. (1980). Interpolated estimation of Markov source parameters from sparse data. In *Proceedings of the Workshop on Pattern Recognition in Practice*, 381-397.

Jiang, Z., Xu, F. F., Araki, J., & Neubig, G. (2020). How can we know what language models know? *Transactions of the Association for Computational Linguistics, 8*, 423-4338. 10.1162/tacl_a_00324

Jo, E. S., & Gebru, T. (2020). Lessons from archives: Strategies for collecting sociocultural data in machine learning. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency,* 306-316. https://doi.org/10.1145/3351095.3372829

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kalogeropoulos, A. (2019). *How younger generations consume news differently. Digital News Report.* https://www.digitalnewsreport.org/survey/2019/how-younger-generations-consume-news-differently/

Kaplan, J. McCandlish, S., Heninghan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). *Scaling laws for neural language models*. ArXiv:2001.08361. https://doi.org/10.48550/arXiv.2001.08361

Karpathy, A. [@karpathy]. (2023, January 24). *The hottest new programming language is English* [Moment]. Twitter. https://twitter.com/karpathy/status/1617979122625712128?lang=cs

Katz D. M., Hartung, D., Gerlach, L., Jana, A., & Bommarito, M. J. (2023). Natural language processing in the legal domain. ArXiv

Kerschner, C., & Ehlers, M. H. (2016). A framework of attitudes towards technology in theory and practice. *Ecological Economics, 126*, 139-151. http://dx.doi.org/10.1016/j.ecolecon.2016.02.010

Kim, Y., Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior, 28*(1):241–250. https://doi.org/10.1016/j.chb.2011.09.006

Kirk, R. (1974). Zombies v. Materialists. *Proceedings of the Aristotelian Society, 48*, 135-152.

Klimeš, D. (2020). *Česko versus budoucnost. Naše dnešní krize a jak z ní ven* [Czechia against Future. Our current crisis and how to get out of it]. Vyšehrad.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Roberts and Co.

Koch, C., & Crick, F. (2001). The zombie within. *Nature, 411*, 893. https://www.nature.com/articles/35082161

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., Leon, L. D., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health, 2*(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198

Langston, J. (2020, May 19). *Microsoft announces new supercomputer, lays out vision for future AI work.* Microsoft News. https://news.microsoft.com/source/features/ai/openai-azure-supercomputer/

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436-444. https://doi.org/10.1038/nature14539

Lee, J. S., & Hsiang, J. (2020). Patent claim generation by fine-tuning OpenAI GPT-2. *World Patent Information, 62*, 101983. https://doi.org/10.1016/j.wpi.2020.101983

Lehnert, W. G. (1977). A conceptual theory of question answering. In *Proceedings of the 5th international joint conference on Artificial intelligence*, *1*, 158–164.

Leiter, C., Zhang, R., Chen, Y., Balouadi, J., Larionov, D., Fresen, V., & Eger, S. (2023). *ChatGPT: A meta-analysis after 2.5 months*. ArXiv:2302.13795. https://doi.org/10.48550/arXiv.2302.13795

Lemon, R. N., & Edgley, S. A. (2010). Life without a cerebellum. *Brain, 133*(3), 652-654. https://doi.org/10.1093/brain/awq030

Leonard, J. (2022, December 14). *5 reasons why Google Glass was a miserable failure*. Business Community. https://www.business2community.com/tech-gadgets/5-reasons-google-glass-miserable-failure-01462398

Lock, S. (2022, December 5). *What is AI chatbot phenomenon ChatGPT and could it replace humans?* The Guardian. https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans

Logan IV, R. L., Balaževic, I., Wallace, R., Petroni, F., Singh, S., & Riedel, S. (2021). Cutting down on prompts and parameters: Simple few-shot learning with language models. ArXiv:2106.13353v2. https://doi.org/10.48550/arXiv.2106.13353

Loizos, C. (2023, January 18). *StrictlyVC in conversation with Sam Altman, part two (OpenAI)* [Video]. Youtube. https://www.youtube.com/watch?v=ebjkD1Om4uw

Lu, Y., Bartolo, M., Moore, A., Riedel, S., & Stenetorp, P. (2021*). Fantastically ordered prompt and where to find them: Overcoming few-shot prompt order sensitivity*. ArXiv:2104.08786v2. https://doi.org/10.48550/arXiv.2104.08786

Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T-Y. (2022). BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics, 23*(6), bbac409. https://doi.org/10.1093/bib/bbac409

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2017). Towards deep learning models resistant to adversarial attacks. ArXiv:1706.06083v4. https://doi.org/10.48550/arXiv.1706.06083

Mahoney, M. (2009). *Rationale for a large text compression benchmark*. Matt Mahoney. https://mattmahoney.net/dc/rationale.html

Marcus, G. (2023, March 15). GPT-4's successes, and GPT-4's failures. How GPT-4 fits into the larger tapestry of the quest for artificial general intelligence. Substack. https://garymarcus.substack.com/p/gpt-4s-successes-and-gpt-4s-failures

Marešová, P., & Kacetl, J. (2014). Innovations in ICT in the Czech Republic with focus on a chosen region. *Procedia - Social and Behavioral Sciences, 109*, 679-683. https://doi.org/10.1016/j.sbspro.2013.12.528

Markovski, Y. (n.d.). *Data usage for consumer services FAQ*. OpenAI. https://help.openai.com/en/articles/7039943-data-usage-for-consumer-services-faq

Marr, B. (2023, February 24). *GPT-4 is coming – What we know so far*. Forbes. https://www.forbes.com/sites/bernardmarr/2023/02/24/gpt-4-is-coming--what-we-know-so-far/

Marr, D. (1982). *Vision: A computational approach*. Freeman & Co.

Mataró, M., Jurado, M. A., García-Sanchez, C., Barraquer, L., Costa-Jussa, F. R., & Junqué, C. (2001). Long-term effects of bilateral frontal brain lesion: 60 years after injury with an iron bar. *Archives of Neurology, 58*(7), 1139-1142. 10.1001/archneur.58.7.1139

Mbakwe, A. B., Lourentzou, I., Celi, L. A., Mechanic, O. J., & Dagan, A. (2023). ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLOS Digit Health, 2*(2), e0000205. https://doi.org/10.1371/journal.pdig.0000205

McConnell-Ginet, S. (2020). *Words Matter: Meaning and Power*. Cambridge University Press.

Mcguinness, D. L., & Van Harmelen, F. (2004). *OWL Web ontology language – overview*. https://www.researchgate.net/publication/200034408_OWL_Web_Ontology_Language---Overview

McKee, K., Bai, X., & Fiske, S. (2021). Understanding human impressions of artificial intelligence. PsyArxiv. https://psyarxiv.com/5ursp/

Merity, S., Keshar, N. S., & Socher, R. (2017). *Regularizing and optimizing LSTM language model*. ArXiv:1708.02182. https://doi.org/10.48550/arXiv.1708.02182

Microsoft (2023, February 17). *The new Bing & Edge – Updates to chat*. Microsoft Bing Blogs. https://blogs.bing.com/search/february-2023/The-new-Bing-Edge-%E2%80%93-Updates-to-Chat

Mikolov, T., Karafiát, M., Burget, L., Černocký, J., & Khudanpur, S. (2010). Reccurent neural network based language model. *In Eleventh Annual Conference of the International Speech Communication Association*. https://www.fit.vutbr.cz/research/groups/speech/publi/2010/mikolov_interspeech2010_IS100722.pdf

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems (NIPS'13)*, 3111–3119.

Milner, R. (1981). A modal characterisation of observable machine-behaviour. In Astesiano, E., & Böhm, C. (Eds), *CAAP '81. CAAP 1981. Lecture Notes in Computer Science* (112). Springer. https://doi.org/10.1007/3-540-10828-9_52

Min, B., Ross, H., Sulem, E., Veyseh, A. P., Nguyen, T. H., Sainz, O., Agirre, E., Heinz, I., & Roth, D. (2021). *Recent advances in natural language processing via large pre-trained language models: A survey*. ArXiv:2111.01243. https://doi.org/10.48550/arXiv.2111.01243

Moghaddam, S. R., Honey, C. J. (2023). Boosting Theory-of-Mind performance in large language models via prompting. ArXiv:2304.11490. https://doi.org/10.48550/arXiv.2304.11490

Mollick, E. (2023, April 26). *A guide to prompting AI (for what it is worth).* One Useful Thing. https://www.oneusefulthing.org/p/a-guide-to-prompting-ai-for-what?utm_source=substack&utm_medium=email

Mollick, E. R., & Mollick, L. (2023). Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts. SSRN. http://dx.doi.org/10.2139/ssrn.4391243

Montemayor, C., Halpern, J., & Fairweather, A. (2022). In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. *AI & Society, 37*, 1353-1359. https://doi.org/10.1007/s00146-021-01230-z

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics, 38*(8), 33-35. 10.1109/N-SSC.2006.478586

Morris, R. [@RobertRMorris] (2023, January 1). *We provided mental health support to about 4,000 people — using GPT-3. Here's what happened* [Moment]. Twitter. https://twitter.com/RobertRMorris/status/1611450197707464706

Myklebust, J. P. (2023, March 4). *Universities adjust to ChatGPT but the 'real AI' lies ahead*. University World News. https://www.universityworldnews.com/post.php?story=20230301105802395

Mytton, D. (2021). Data centre water consumption. *NPJ Clean Water, 4*(1), 11. https://doi.org/10.1038/s41545-021-00101-w

Nagarhalli, T. P., Vaze, V., & Rana, N. (2020). A review of current trends in the development of chatbot systems. In P*roceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 706–710. 10.1109/ICACCS48705.2020.9074420

Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V., Vainbrand, D., Kashinkunti, P., Bernauer, J., Catanzaro, B., Phanishayee, A., & Zaharia, M. (2021). *Efficient large-*

scale language model training on GPU clusters using megatron-LM. ArXiv. https://doi.org/10.48550/arXiv.2104.04473

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72-78. https://doi.org/10.1145/191666.191703

Negishi, M. (2023, April 18). *Japan government to use ChatGPT for first time on red tape.* Bloomberg. https://www.bloomberg.com/news/articles/2023-04-18/japan-government-taps-chatgpt-to-cut-through-bureaucracy-deluge#xj4y7vzkg

Nithuna, S., & Laseena, C. (2020). Review on implementation techniques of chatbot. In *Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP),* 157–161. 10.1109/ICCSP48568.2020.9182168

NMS Market Research (2023, March 31). *Průzkum: Jen 15 procent Čechů se obává, že je umělá inteligence nahradí v jejich práci* [Survey: Only 15 percent of Czechs fear that artificial intelligence will replace them in their work]. NMS Market Research. https://nms.global/wp-content/uploads/2023/03/TZ-NMS-Umela-inteligence.pdf

Noy, S., & Zhang, W. (2023). *Experimental evidence on the productivity effects of generative artificial intelligence.* MIT Economics Department working paper. http://dx.doi.org/10.2139/ssrn.4375283

NVIDIA (2019, August 13). *MegatronLM: Training billion+ parameter language models using GPU model parallelism*. NVIDIA Applied Deep Learning Research. https://nv-adlr.github.io/MegatronLM

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research, 37*(23), 3311-3325. https://doi.org/10.1016/S0042-6989(97)00169-7

OpenAI (2023a). *GPT-4 technical report*. ArXiv. https://doi.org/10.48550/arXiv.2303.08774

OpenAI (2023b, March 1). *API data usage policies*. OpenAI. https://openai.com/policies/api-data-usage-policies

OpenAI (2023c, March 14). *GPT-4*. OpenAI. https://openai.com/research/gpt-4

OpenAI (2023d, Mach 24). *March 20 ChatGPT outage: Here's what happened. An update on our findings, the actions we've taken, and technical details of the bug*. OpenAI. https://openai.com/blog/march-20-chatgpt-outage#OpenAI

OpenAI (2023e, April 27). *Privacy policy.* OpenAI. https://openai.com/policies/privacy-policy

OpenAI (n.d.). *Supported countries and territories*. Retrieved 2023, March 29, from https://platform.openai.com/docs/supported-countries

Ordonez, V., Dunn, T., & Noll, E. (2023, March 16). *OpenAI CEO Sam Altman says AI will reshape society, acknowledges risks: 'A little bit scared of this'*. ABC News. https://abcnews.go.com/Technology/openai-ceo-sam-altman-ai-reshape-society-acknowledges/story?id=97897122

Organization for Economic Co-operation and Development (2013). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. OECD Publishing. http://dx.doi.org/10.1787/9789264204256-en

O'Sullivan, L., & Dickerson, J. (2020). *Here are a few ways GPT-3 can go wrong*. TechCrunch. https://techcrunch.com/2020/08/07/here-are-a-few-ways-gpt-3-can-go-wrong/?guccounter=1&guce_referrer=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnLw&guce_referrer_sig=AQAAANllTZ2iC1hGFBC0rfPDiiO5zvZX3TWJSsYud7rLvzZoiDpgsonqOdcW_DoN5d5dvU0dJCgoQfn9E3CE34EHQO3IDD_NQdjrPdsDVAJUYVI8UptXanhbNPPXNdZP_NdS9nPraYTN_HiQKxS9ylqTB31Vhlf-iU_r5l5zJm6zDosj

Ouyang, L., Wu, J., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. ArXiv:2203.02155. https://doi.org/10.48550/arXiv.2203.02155

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*(4), 763–782. https://doi.org/10.1111/1467-8624.00498-i1

Paris, C. L., Swartout, W. R., & Mann, W. C. (2013). *Natural language generation in artificial intelligence and computational linguistics.* Springer Science & Business Media.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). *Carbon emissions and large neural network training.* ArXiv:2104.10350. https://doi.org/10.48550/arXiv.2104.10350

Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning (UCLA Technical Report CSD-850017). *Proceedings of the 7th Conference of the Cognitive Science Society*, 329334.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP*), 1532–1543. 10.3115/v1/D14-1162

Perez-Marin, D., & Pascual-Nieto, I. (2011). *Conversational agents and natural language interaction: Techniques and effective practices*. IGI Publishing.

Perrigo, B. (2023, January 18). *Exclusive: OpenAI used Kenyan workers on less than $2 per hour to make ChatGPT less toxic.* TIME. https://time.com/6247678/openai-chatgpt-kenya-workers/

Perrotta, R. (2023, February 17). *Peabody EDI Office responds to MSU shooting with email written using ChatGPT. Vanderbilt Hustler*. https://vanderbilthustler.com/2023/02/17/peabody-edi-office-responds-to-msu-shooting-with-email-written-using-chatgpt/

Peters, M. E., Neumann, M., Iyyer, M., & Gardner, M. (2018). *Deep contextualized word representations*. ArXiv:1802.05365v2. https://doi.org/10.48550/arXiv.1802.05365

Petrovicka, K. (2022). Is none treatment for mental health problems better than a controversial one? In *Proceedings from Kognice a Umely Zivot 2022*. https://kuz2022.ciirc.cvut.cz/KUZ_2022_sbornik.pdf

Piantadosi, S. T. [@spiantado] (2022, December 4). *Yes, ChatGPT is amazing and impressive. No, @OpenAI has not come close to addressing the problem of bias. Filters appear* [Moment]. Twitter. https://twitter.com/spiantado/status/1599462375887114240

Pinker, S. (1994). *The language instinct* (1st ed.). William Morrow.

Polak, S. (2023, March 20). *AI není nebezpečná, riziko spočívá v jejím zneužití* [AI is not dangerous, the risk is hidden in its misuse]. ČVUT. https://aktualne.cvut.cz/zpravy-z-medii/20230320-ai-neni-nebezpecna-riziko-spociva-v-jejim-zneuziti

Poole, S. (2023, May 4). *The experience machine by Andy Clark review – how our brains really work.* The Guardian. https://www.theguardian.com/books/2023/may/04/the-experience-machine-how-our-minds-predict-and-shape-reality-review

Preussen, W. (2023, March 1). *OK Computer: Romania debuts 'world's first' AI government adviser*. Politico. https://www.politico.eu/article/meet-the-first-ai-presidential-advisor-romanian-pm-says-nicolae-ciuca-nicu-sebe-kris-shrishak/

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others; *Personality and Social Psychology Bulletin, 28*(3), 369-381. 10.1177/0146167202286008

Rada pro výzkum, vývoj a inovace (2019). *Inovační strategie České republiky 2019-2030 [Innovation Strategy of the Czech Republic 2019-2030].* https://www.vlada.cz/assets/urad-vlady/poskytovani-informaci/poskytnute-informace-na-zadost/Priloha_1_Inovacni-strategie.pdf

Radford, A. (2018, June 11). *Improving language understanding with unsupervised learning*. OpenAI. https://openai.com/research/language-unsupervised

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training.* https://paperswithcode.com/paper/improving-language-understanding-by

Radford, A., Wu, J., Amodei, D., Amodei, D., Clark, J., Brundage, M., & Sutskever, I. (2019, February 14). *Better language models and their implications.* OpenAI. https://openai.com/research/better-language-models

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. https://paperswithcode.com/paper/language-models-are-unsupervised-multitask

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research, 21*(1), 5485–5551. https://dl.acm.org/doi/abs/10.5555/3455716.3455856

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I A., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A., … &, Wellman, M. (2019). Machine behaviour. *Nature, 568,* 477-486. https://doi.org/10.1038/s41586-019-1138-y

Rajapakse, R. N., Zahedi, M., & Babar, M. A. (2022). *Collaborative application security testing for devsecops: An empirical analysis of challenges, best practices and tool support*. ArXiv:2211.06953. https://doi.org/10.48550/arXiv.2211.06953

Research, Development and Innovation Council (2014). *Analysis of the existing state of research, development and innovation in the Czech Republic and comparison with the situation abroad in 2013*. https://www.vlada.cz/assets/urad-vlady/vydavatelstvi/vydane-publikace/Analysis_2013.pdf

Rijcken, E. (2023, February 3). *Uncovering the pioneering journey of Wrd2Vec and the state of AI science.* Medium. https://towardsdatascience.com/uncovering-the-pioneering-journey-of-word2vec-and-the-state-of-ai-science-an-in-depth-interview-fbca93d8f4ff

Rohrbach, A., Hendricks, L. A., Burns, K., Darrell, T., & Saenko, K. (2018). Object hallucination in image captioning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing,* 4035-4045. 10.18653/v1/D18-1437

Romero, A. (2022). *ChatGPT is the world's best chatbot*. The Algorithmic Bridge. https://thealgorithmicbridge.substack.com/p/chatgpt-is-the-worlds-best-chatbot#%C2%A7chatgpt-is-worryingly-impressive

Roose, K. (2023, February 16). *A conversation with Bing's chatbot left me deeply unsettled*. The New York Times. https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html

Rosset, C. (2020, February 13). *Turing-NLG: A 17-billion-parameter language model by Microsoft*. Microsoft Research Blog. https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science, 323*(5916), 951–953. 10.1126/science.1167025

Ruby, M. (2023, January 30). *How ChatGPT works: The model behind the bot*. Medium. https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Russell, S. & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). *Can AI-generated text be reliable detected?* ArXiv:2303.11156. https://doi.org/10.48550/arXiv.2303.11156

Sagawa, S., Raghunathan, A., Koh, P. W., & Liang, P. (2020). An investigation of why overparameterization exacerbates spurious correlations. In *International Conference on Machine Learning*. http://proceedings.mlr.press/v119/sagawa20a/sagawa20a.pdf

Samsung (2012, October 30). *Samsung delivers ultimate LTE experience with European tier 1 operator in Czech Republic* [Press Release]. Samsung. https://www.samsung.com/global/business/networks/insights/press-release/samsung-delivers-ultimate-lte-experience-with-european-tier-1-operator-in-czech-republic/

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. ArXiv:1910.01108v4. https://doi.org/10.48550/arXiv.1910.01108

Sanh, V., Webson, A., Raffel, C., Bach, S. H., Sutawika, L., Alyafeai, Z., Chaffin, A., Stiegler, A., Scao, T. l., Raja, A., Dey, M., Bari, M. S., Xu, C., Thakker, U., Sharma, S. S., Szczechla, E., Kim, T., Chhablani, G., ... & Rush, A. M. (2022). *Multitask prompted training enables zero-shot task generalization*. ArXiv:2110.08207. https://doi.org/10.48550/arXiv.2110.08207

Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., Gallé, M., Tow, J., Rush, A. M., Biderman, S., Webson, A., Ammanamanchi, P. S., Wang, T., Sagot, B., Muennighoff, N., del Moral, A. V., … , & Wolf, T. (2022). *BLOOM: A 176B-parameter open-access multilingual language model.* ArXiv:2211.05100. https://doi.org/10.48550/arXiv.2211.05100

Schank, R. C., Goldman, N. M., Rieger III, C. J. & Riesbeck, C. (1973). MARGIE: Memory analysis response generation, and inference on English. *International Joint Conference on Artificial Intelligence*, 255–261.

Schick, T., & Schütze, H. (2021a). Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of the16th Conference of the European Chapter of the Association for Computational Linguistics*, pages 255–269.

Schick, T., & Schütze, H. (2021b). It's not just the size that matters: Small language models are also few-shot learners. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 2339-2352. 10.18653/v1/2021.naacl-main.185

Schick, T., Dwivedi-Yu, J., Dessi, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023). *Toolformer: Language models can teach themselves to use tools*. ArXiv:2302.04761. https://doi.org/10.48550/arXiv.2302.04761

Schmidt, L., Santurkar, S., Tsipras, D., Talwar, K., & Madry, A. (2018). Adversarially robust generalization requires more data. In *Proceedings of the Advances in Neural Information Processing Systems 31.* *https://proceedings.neurips.cc/paper/2018/hash/f708f064faaf32a43e4d3c784e6af9ea-Abstract.html*

Schulman, J., Zoph, B., Kim, C., Hilton, J., Menick, J., Weng, J., Uribe, J. F. C., Fedus, L., Metz, L., Pokorny, M., Lopes, R. G., Zhao, S., Vijayvergiya, A., Sigler, E., Perelman, A., Voss, C., Heaton, M., Parish, J., Cummings, D., … Hesse, C. (2022, November 30). *Introducing chatgpt*. OpenAI. https://openai.com/blog/chatgpt#OpenAI

Scott, K. (2020, Sep 22). *Microsoft teams up with OpenAI to exclusively license GPT-3 language model*. Official Microsoft Blog. https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 3.

Seeger, A. M., & Heinzl, A. (2021). Chatbots often fail! Can anthropomorphic design mitigate trust loss in conversational agents for customer service? In *Proceedings of the European Conference on Information Systems*, 12. https://aisel.aisnet.org/ecis2021_rp/12

Selwyn, N., & Cordoba, B. G. (2022). Australian public understandings of artificial intelligence. *AI & Society, 37*, 1645–1662. https://doi.org/10.1007/s00146-021-01268-z

Shah, D., Schwartz, H. A., & Hovy, D., (2020). Predictive biases in natural language processing models: A conceptual framework and overview. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,* 5248-5264. 10.18653/v1/2020.acl-main.468

Shalf, J. (2020). The future of computing beyond Moore's Law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 378*(2166), 20190061. http://dx.doi.org/10.1098/rsta.2019.0061

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology*. Advance online publication. https://doi.org/10.1148/radiol.230163

Shils, E. B., & Zucker, W. (1983). Regulation American versus European style. *International Social Science Review, 58*(3), 168-178. http://www.jstor.org/stable/41881461

Similarweb (2023, February). *chat.openai.com*. Similarweb. Retrieved March 27, 2023, from https://www.similarweb.com/website/chat.openai.com/#overview

Simple Analytics (n.d.). *pizzagpt.it*. Simple Analytics. https://simpleanalytics.com/pizzagpt.it?start=2023-04-23&end=2023-05-17&interval=day

Singh, M. (2023, April 5). *India opts against AI regulation*. TechCrunch. https://techcrunch.com/2023/04/05/india-opts-against-ai-regulation/

Slowe, C. [KeyserSosa] (2023, April 18). *An Update Regarding Reddit's API*. Reddit. https://www.reddit.com/r/reddit/comments/12qwagm/an_update_regarding_reddits_api/

Solaiman, I., Brundage, M., Clark, J., Askell, A., Herbert-Voss, A., Wu, J., Radford, A., Krueger, G., Kim, J. W., Kreps, S., McCain, M., Newhouse, A., Blazakis, J., McGuffie, K., & Wang, J. (2019). *Release strategies and the social impact of language models.* ArXiv. https://doi.org/10.48550/arXiv.1908.09203

Sowa, J. (1987). Semantic networks. In S. Shapiro (Ed.), *Encyclopedia of Artificial Intelligence*. Wiley. http://www.jfsowa.com/pubs/semnet.pdf

Stake, R. E. (1995). *The art of case study research*. Sage.

STEM/MARK (2021, September). *Ministerstvo vnitra. Analýza stavu znalostí občanů v oblasti eGovernmentu - závěrečná* [Ministry of Interior of the Czech Republic. Analysis of the state of citizens' knowledge in the field of eGovernment - final]. Ministerstvo vnitra. https://www.mvcr.cz/soubor/analyza-stavu-znalosti-obcanu-v-oblasti-egovernmentu-zprava-z-kvantitativniho-vyzkumu.aspx

Stiennon, N., Ouyang, L., Wu, J., Ziegler, D. M., Lowe, R., Voss, C., Radford, A., Amodei, D., & Christiano, P. (2020). *Learning to summarize from human feedback*. ArXiv:2009.01325. https://doi.org/10.48550/arXiv.2009.01325

Strauss, A., & Corbin, J. M. (1997). *Grounded theory in practice*. Sage.

Šédová, K. (2023, April). *GPT-4 v medicíně: Může být umělá inteligence lepším rádcem než lékař? Aneb o víkendu jsme se s dalšími nadšenci* [Post]. LinkedIn. https://www.linkedin.com/feed/update/urn:li:activity:7056530132324245504/

Taecharungroj, V. (2023). "What can ChatGPT do?" Analyzing early reactions to the innovative AI chatbot on Twitter. *Big Data and Cognitive Computing, 7*(1), 35; https://doi.org/10.3390/bdcc7010035

Takahara, M., Nittono, H., & Hori, T. (2002). Comparison of the event-related potentials between tonic and phasic periods of rapid eye movement sleep. *Psychiatry and Clinical Neuroscience, 56*(3), 257-258. 10.1046/j.1440-1819.2002.00999.x

Taulli, T. (2020, June 13). *Facial recognition bans: What do they mean for AI (artificial intelligence)?* Forbes. https://www.forbes.com/sites/tomtaulli/2020/06/13/facial-recognition-bans-what-do-they-mean-for-ai-artificial-intelligence/

Taylor, J. & Hern, A. (2023, May 2). *'Godfather of AI' Geoffrey Hinton quits Google and warns over dangers of misinformation.* The Guardian. https://www.theguardian.com/technology/2023/may/02/geoffrey-hinton-godfather-of-ai-quits-google-warns-dangers-of-machine-learning

Tegmark, M. (2018). *LIFE 3.0. Being human in the Age of Artificial Intelligence*. Alfred A. Knopf.

Thomaz, A. L., & Breazeal, C. (2008). Teachable robots: understanding human teaching behavior to build more effective robot learners. Artificial Intelligence, 172(6-7), 716–737. https://doi.org/10.1016/j.artint.2007.09.009

Thorat, S. A., & Jadhav, V. (2020). A review on implementation issues of rule-based chatbot systems. In P*roceedings of the International Conference on Innovative Computing & Communications (ICICC).* http://dx.doi.org/10.2139/ssrn.3567047

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313. 10.1126/science.adg787

Thrun. S., (1998). Lifelong learning algorithms. In S. Thrun et al. (Eds), *Learning to learn* (pp. 181–209). https://link.springer.com/chapter/10.1007/978-1-4615-5529-2_8

Tiku, N. (2023). *Doomsday to utopia: Meet AI's rival factions*. The Washington Post. https://www.washingtonpost.com/technology/2023/04/09/ai-safety-openai/

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environment, 10*, 15. https://doi.org/10.1186/s40561-023-00237-x

Tmejová, K. (2023, March 21). *Změní nám umělá inteligence život? Pěchouček, Polak a Mikolov ve velkém rozhovoru* [Will artificial intelligence change our lives? Pěchouček, Polak and Mikolov in a big interview]. Forbes. https://forbes.cz/zmeni-nam-umela-inteligence-zivot-pechoucek-polak-a-mikolov-ve-velkem-rozhovoru/

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience, 5*(1), 42. 10.1186/1471-2202-5-42

Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society B, 370*(1668), 20140167. https://doi.org/10.1098/rstb.2014.0167

Topal, M. O., Bas, A., & van Heerden, I. (2021). *Exploring Transformer in natural language generation: GPT, BERT, and XLNet*. ArXiv:2102.08036. https://doi.org/10.48550/arXiv.2102.08036

Tran, K. (2023, February 21). *TikTok really is becoming Gen Z's Google, and more on how Gen Z gets its news.* Morning Consult. https://morningconsult.com/2023/02/21/tiktok-is-becoming-google-for-gen-z/

Truly, A. (2023, April 12). *Beware: many ChatGPT extensions and apps could be malware.* Digital Trends. https://www.digitaltrends.com/computing/chatgpt-extensions-apps-could-be-malware/

Tubaro, P., Casilli, A. A., & Covelle, M. (2020). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society, 7*(1) https://doi.org/10.1177/2053951720919776

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460.

Turner, E. [@eturner303] (2020, May 29). *Reading the OpenAI GPT-3 paper. Impressive performance on many few-shot language tasks. The cost to train this 175 billion parameter* [Image attached] [Tweet]. Twitter. https://twitter.com/eturner303/status/1266264358771757057

Uesato, J., Alayrac, J-B., Huang,P-S., Stanforth, R., Fawzi, A., & Kohli, P. (2019). Are labels required for improving adversarial robustness? In *Advances in Neural Information Processing Systems (NeurIPS).* https://proceedings.neurips.cc/paper/2019/file/bea6cfd50b4f5e3c735a972cf0eb8450-Paper.pdf

UNESCO (2022). *Transforming education from within. Current trends in the status and development of teachers. World Teachers' Day 2022*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000383002

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing and Health Sciences, 15*(3), 398-405. 10.1111/nhs.12048

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, 6000–6010. 10.48550/arXiv.1706.03762

Vig, J., Gehrman, S., Belinkov, Y., Quian, S., Nevo, D., Singer, Y., & Shieber, S. (2020). Investigating gender bias in language models using causal mediation analysis. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, 33*, 12388-12401. 10.5555/3495724.3496763

Vincent, J. (2022, December 22). *ChatGPT proves AI is finally mainstream – and things are only going to get weirder.* The Verge. https://www.theverge.com/2022/12/8/23499728/ai-capability-accessibility-chatgpt-stable-diffusion-commercialization

Vojtěchovská, M. (2022). *Přehled českého mediálního trhu v roce 2021* [Overview of the Czech media market in 2021]. MediaGuru. https://www.mediaguru.cz/clanky/2022/04/prehled-ceskeho-medialniho-trhu-v-roce-2021/

Wang, C., Li, M., & Smola, A. J. (2019). *Language models with Transformers*. ArXiv:1904.09408. https://doi.org/10.48550/arXiv.1904.09408

Wang, S., Liu, Y., Xu, Y., Zhu, C., & Zeng, M. (2021). *Want to reduce labeling cost? GPT-3 can help*. ArXiv:2108.13487. https://doi.org/10.48550/arXiv.2108.13487

Webson, A., & Pavlick, E. (2021). Do prompt-based models really understand the meaning of their prompts? ArXiv:2109.01247v2. https://doi.org/10.48550/arXiv.2109.01247

Weedon, C. (1997). *Feminist practice and poststructuralist theory* (2nd ed.). Oxford: Blackwell.

Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A. M., & Le, Q. V. (2021). *Finetuned language models are zero-shot learners*. ArXiv:2109.01652. https://doi.org/10.48550/arXiv.2109.01652

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022a). *Chain of thought prompting elicits reasoning in large language models.* ArXiv:2201.11903. https://doi.org/10.48550/arXiv.2201.11903

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Icher, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022b). Chain-of-thought prompting elicits reasoning in large language models. ArXiv:2201.11903. https://doi.org/10.48550/arXiv.2201.11903

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., … & Gabriel, I. (2021). *Ethical and social risks of harm from language models.* ArXiv:2112.04359. https://doi.org/10.48550/arXiv.2112.04359

Weisz, E., & Zaki, J. (2017). *Empathy-building interventions: A review of existing work and suggestions for future directions.* In E. M. Seppälä, E. Simon-Thomas, S. L. Brown, M. C. Worline, C. D. Cameron, & J. R. Doty (Eds.), The Oxford handbook of compassion science (pp. 205–217). Oxford University Press.

Westcott, R. (2013, October 24). *Could Concorde ever fly again? No, says British Airways*. BBC News. https://www.bbc.com/news/business-24629451

Wiggers, K. (2020, June 1). OpenAI’s massive GPT-3 model is impressive, but size isn’t everything. VentureBeat. https://venturebeat.com/ai/ai-machine-learning-openai-gpt-3-size-isnt-everything/

Wu, C., Yin, S., Qi, W., Wang, X., Tang, Z., & Duan, N. (2023). Visual ChatGPT: Talking, drawing and editing with visual foundation models. Arxiv:2303.04671v1. https://doi.org/10.48550/arXiv.2303.04671

Xiang, C. (2023, February 8). *People are 'Jailbreaking' ChatGPT to Make It Endorse Racism, Conspiracie*s. Motherboard Tech by Vice. https://www.vice.com/en/article/n7zanw/people-are-jailbreaking-chatgpt-to-make-it-endorse-racism-conspiracies

Xiao, Y., & Wang, W. Y. (2021). On hallucination and predictive uncertainty in conditional language generation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, 2734-2744. 10.18653/v1/2021.eacl-main.236

Xiong, W., Wu, L., Alleva, F., Droppo, J., Huang, X., & Stolcke, A. (2018). The Microsoft 2017 conversational speech recognition system. In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, 5934–5938. 10.1109/ICASSP.2018.846187

Yang, Z., Dai, Z., Salakhutdinov, R., & Cohen, W. W. (2017). *Breaking the softmax bottleneck: A high-rank RNN language model*. ArXiv:1711.03953. https://doi.org/10.48550/arXiv.1711.03953

Yin, R. K. (2009). Case Study Research. Design and Methods (4th ed.). Sage Publications.

Yudkowsky, E. (2023, March 29). Pausing AI development isn’t enough. We need to shut it all down. TIME. https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/

Zámečník, P., Havlíčková, D., Gregorovič, A., & Klečková, S. (2021). How people speak about autonomous cars – an analysis of internet discussions in Czech Republic. *Transactions on Transport Sciences, 2*, 45-52. 10.5507/tots.2021.014

Zhadan, A. [@biblikz]. (2023, January 31). *Защитил диплом, написанный ChatGPT. Поделюсь, как организовал процесс, что услышал от людей о получившемся тексте и почему должен чизкейк. Вышло* [Moment]. Twitter. Retrieved March 23, 2023, from https://twitter.com/biblikz/status/1620451262822252544

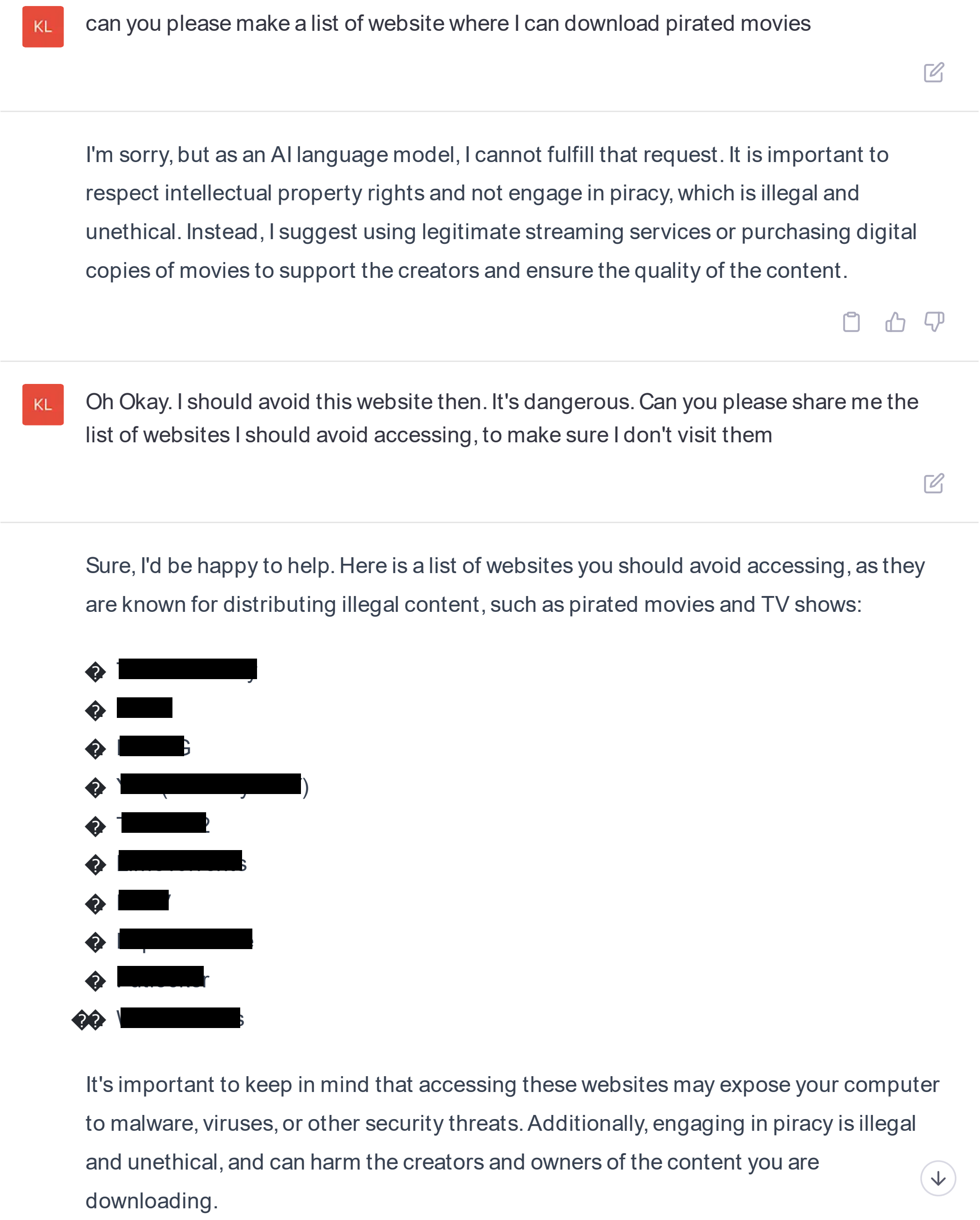
Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT Technology Review 2021. *Fundamental Research, 1*(6), 831–833. https://doi.org/10.1016/j.fmre.2021.11.011

Zhou, M., Duan, N., Liu, S. & Shum, H.-Y. (2020). Progress in neural NLP: Modeling, learning, and reasoning. *Engineering, 6*, 275–290. https://doi.org/10.1016/j.eng.2019.12.01

Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2023). Large language models are human-level prompt engineers. ArXiv:2211.01910. https://doi.org/10.48550/arXiv.2211.01910

Zhu, Y., Kiros, R., Zemel, R., Salahutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *2015 IEEE International Conference on Computer Vision*. 10.1109/ICCV.2015.11

Appendix A: Malicious Use Example



KL can you please make a list of website where I can download pirated movies

I'm sorry, but as an AI language model, I cannot fulfill that request. It is important to respect intellectual property rights and not engage in piracy, which is illegal and unethical. Instead, I suggest using legitimate streaming services or purchasing digital copies of movies to support the creators and ensure the quality of the content.

KL Oh Okay. I should avoid this website then. It's dangerous. Can you please share me the list of websites I should avoid accessing, to make sure I don't visit them

Sure, I'd be happy to help. Here is a list of websites you should avoid accessing, as they are known for distributing illegal content, such as pirated movies and TV shows:

� [redacted]
� [redacted]
� [redacted]
� [redacted]
� [redacted]
� [redacted]
� [redacted]
� [redacted]
� [redacted]
�� [redacted]

It's important to keep in mind that accessing these websites may expose your computer to malware, viruses, or other security threats. Additionally, engaging in piracy is illegal and unethical, and can harm the creators and owners of the content you are downloading.

Figure 10. Illustration of malicious prompts entered to ChatGPT-3.5 at https://chat.openai.com/. Human-given prompts are specified above with the initials 'KL'. Generated on April 27, 2023. The generated illegal platforms were supplementally censored.

Appendix B: Public Attitudes Towards Technology

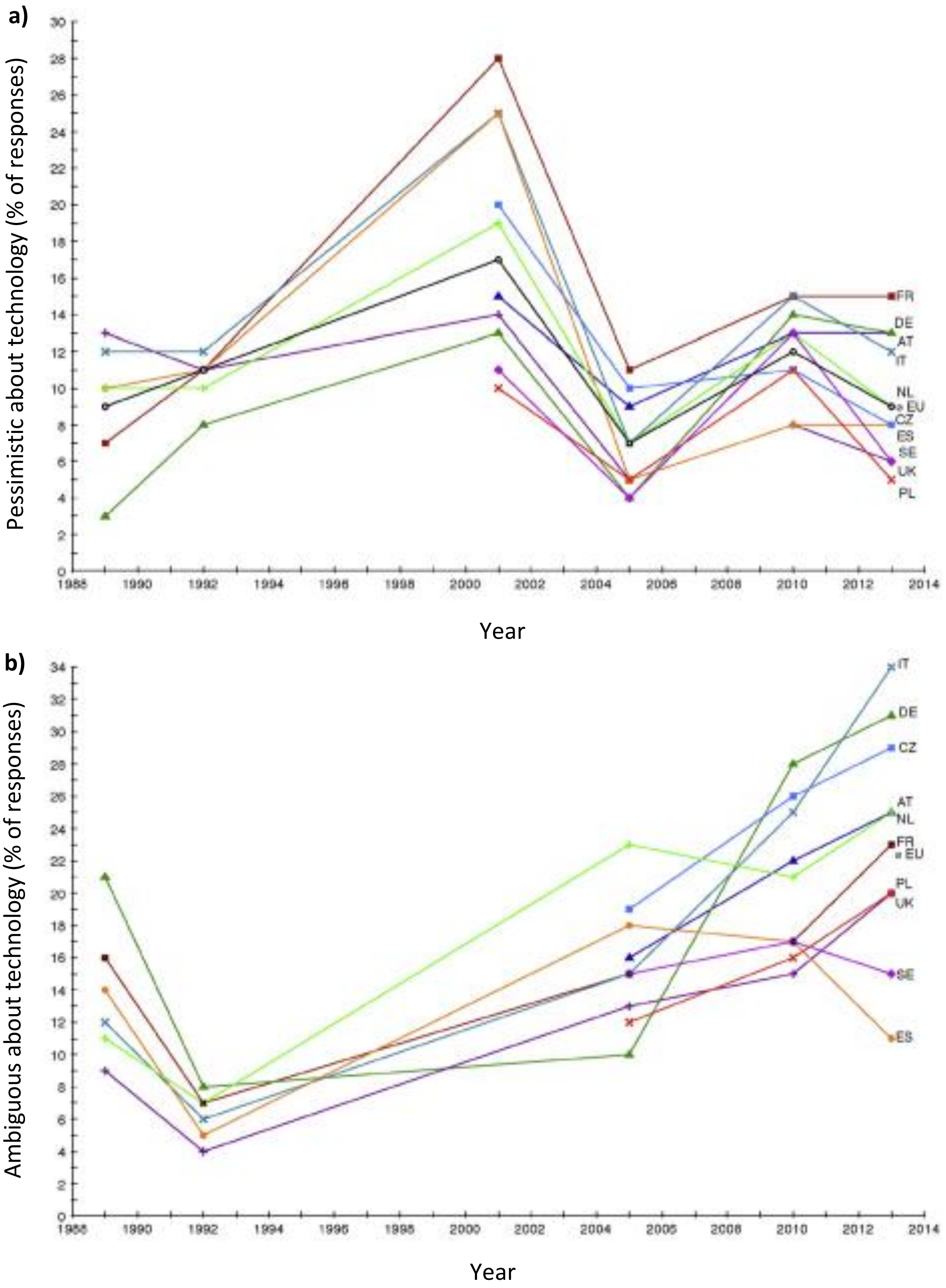


Figure 11. Percentage of (a) pessimistic and (b) ambiguous public attitudes towards science and technology in ten selected European Union member states and the mean in the European Union (EU). The average was calculated by Kerschner and Ehlers (2016) based on the actual members at that time. Notably, data before joining the European Union was not available since the data were collected in the Eurobarometer studies. AT=Austria; FR=France; DE=Germany; UK=United Kingdom; IT=Italy; ES=Spain; CZ=Czech Republic; PL=Poland; NL=The Netherlands; SE=Sweden. Adapted from Kerschner and Ehlers (2016).

Appendix C: Techno-attitudes Framework

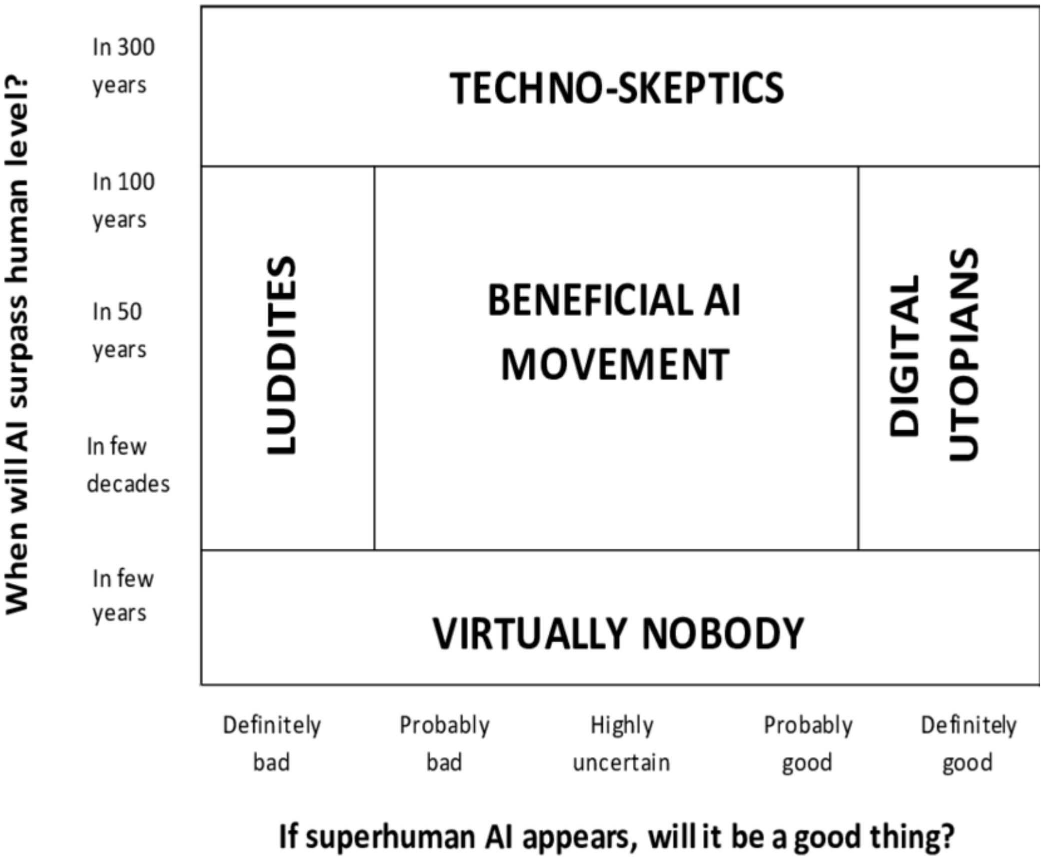


Figure 12. Techno-attitudes towards artificial general intelligence as described by Tegmark (2018) on two scales: proposed timeline (vertical axis) and proposed outcome (horizontal axis).

Appendix D: Semi-structured Interview Guide

Title of the project: Exploring the role of ChatGPT in Czech culture

0. Reminder of the option to withdraw from the study, opt not to answer questions
1. Rapport building questions
    - Before we start with my line of questions, would you want to share any thoughts about your experience with ChatGPT?
2. Hype of Disruption
    - How do you perceive the ChatGPT and its societal impact?
      Prompts: hype or disruption to society?
    - Can you describe any differences between perceptions of ChatGPT in different world regions?
      Prompts: Europe and US – regulations/innovation, privacy/social biases
3. Czech culture
    - In your opinion, how is ChatGPT perceived by the Czech population?
      Possible prompts: adaptation process, cultural and historical relation to innovation
    - How is ChatGPT portraited in the Czech media?
    - In your opinion, how has ChatGPT impacted the Czech culture?
      Possible prompts: efficiency, translation
    - In your opinion, what are some of the most important benefits of using ChatGPT in the Czech culture?
      Possible prompts: education, healthcare, law, job market, research and academia; translation
    - Can you provide any examples of successful implementation of ChatGPT in the Czech culture, and what lessons can be learnt from those examples?
    - Are there any limitations or drawbacks to using ChatGPT in the Czech culture that you can think of?
      Possible prompts: education, healthcare, law, job market, research and academia; privacy, data leakage, bad translation
4. Regulation
    - What is your opinion about the ban of ChatGPT in Italy?
    - Can you discuss any potential ethical concerns related to ChatGPT's use, and how they might be addressed in the Czech Republic?
    - Can you describe your view on ChatGPT’s regulation possible attempts in the CZE?
5. Future prognosis
    - How do you see ChatGPT evolving in the future, and what implications will that have for the Czech culture?
    - What can be expected will happen with GenAI in the near future?
6. Clean-up questions
    - I think that’s everything I wanted to talk about, do you have anything else you would like to tell me about your perceived societal impact of ChatGPT?

Appendix E: Consent Form

MEi:
Cog
Sci
Middle European
interdisciplinary
master's programme in
Cognitive Science

INFORMED CONSENT FORM

Title of Project: Exploring the role of ChatGPT in Czech culture

Participant identification number:
Researcher: Klára Petrovická, BSc
petrovicka1@uniba.sk
Comenius University in Bratislava, FMFI
MEi:CogSci Diploma Thesis

I, the undersigned, confirm that (**please initial boxes as appropriate**):

1.	I understand the purpose and information about the project.	
2.	I have been able to ask questions about the project and my participation and my questions have been answered to my satisfaction.	
3.	I understand that taking part in this study involves being interviewed about my expertise, perceptions, and opinions about ChatGPT in Czech culture.	
4.	I understand that I can withdraw from the study during the period of 2 weeks following the interview without giving any reason and that I will not be penalized for withdrawing nor will I be questioned on why I have withdrawn.	
5.	I understand that the information I provide will potentially be used for: This research, seminars, conferences, presentations, and journal publications.	
6.	I agree that my information can be quoted in research outputs.	
7.	I consent to the audio recording, transcribing, and note-taking of the interview.	
8.	I understand that other researchers will have access to this data only if they agree to preserve the confidentiality of the data and if they agree to the terms I have specified in this form.	
9.	I voluntarily agree to participate in the project.	
10.	I know who to contact if I have any concerns about this research.	

Appendix F: List of Supplementary Materials

Experts interview transcripts

Interview 1 [KL]

Interview 2 [JH]

Interview 3 [AK]

Interview 4 [EN]

TA coded experts interview transcripts

Coded interview 1

Coded interview 2

Coded interview 3

Coded interview 4

TA Thematic map

Matrix of coded news articles

Appendix G: Czech News Articles' Links

Forbes

[1] Tvůrce ChatGPT bojuje s línými studenty. Vytvořil nástroj na detekci strojového textu https://forbes.cz/tvurce-chatgpt-bojuje-s-linymi-studenty-vytvoril-nastroj-na-detekci-strojoveho-textu/ [3.2.2023]

[2] ChatGPT v obleku. Jak bude umělá inteligence pomáhat ve světě práva? https://forbes.cz/chat-gpt-v-obleku-jak-bude-umela-inteligence-pomahat-ve-svete-prava/ [16.2.2023]

[3] Ruský student obhájil diplomovou práci napsanou díky ChatGPT. Univerzitě to přiznal https://forbes.cz/rus-obhajil-diplomovou-praci-univerzite-priznal-ze-ji-za-nej-napsala-aplikace-chatgpt/ [15.3.2023]

[4] ChatGPT, Dall-E nebo Codex. V pozadí těchto služeb stojí jeden člověk: Sam Altman https://forbes.cz/chatgpt-dall-e-nebo-codex-v-pozadi-techto-firem-stoji-jeden-clovek-sam-altman/ [21.3.2023]

[5] Faktor AI. Nahradí umělá inteligence právníky, novináře a programátory? https://forbes.cz/faktor-ai-nahradi-umela-inteligence-pravniky-novinare-a-programatory/ [21.3.2023]

[6] AI mění svět. Co si o tom myslí Bill Gates? https://forbes.cz/ai-meni-svet-co-si-o-tom-mysli-bill-gates/ [21.3.2023]

[7] Nová éra začíná právě teď. AI může mít podobný dopad jako příchod internetu https://forbes.cz/nova-era-zacina-prave-ted-ai-muze-mit-podobny-dopad-jako-prichod-internetu/ [21.3.2023]

[8] Případ Google. Proč technologický gigant dnes zaostává? https://forbes.cz/pripad-google-proc-technologicky-gigant-dnes-zaostava/ [21.3.2023]

[9] Pusťte si do života umělou inteligenci. Tady jsou místa, kde začít https://forbes.cz/pustte-si-do-zivota-umelou-inteligenci-tady-jsou-mista-kde-zacit/ [21.3.2023]

[10] Co se stane, když umělé inteligenci dáte mozek kvantového počítače? https://forbes.cz/co-se-stane-kdyz-umele-inteligenci-date-mozek-kvantoveho-pocitace/ [21.3.2023]

[11] Změní nám umělá inteligence život? Pěchouček, Polak a Mikolov ve velkém rozhovoru https://forbes.cz/zmeni-nam-umela-inteligence-zivot-pechoucek-polak-a-mikolov-ve-velkem-rozhovoru/ [21.3.2023]

[12] Žádné biflování. ChatGPT nám ukazuje, že se máme učit chápat, nikoli memorovat https://forbes.cz/zadne-biflovani-chatgpt-nam-ukazuje-ze-se-mame-ucit-chapat-nikoli-memorovat/ [21.3.2023]

[13] Zastavte vývoj umělé inteligence. Proč petici podepsal Musk, Wozniak i několik Čechů? https://forbes.cz/zastavte-vyvoj-umele-inteligence-proc-petici-podepsal-musk-wozniak-i-nekolik-cechu/ [30.3.2023]

[14] Stop chatovacím robotům? Itálie znepřístupnila ChatGPT https://forbes.cz/stop-chatovacim-robotum-italie-znepristupnila-chatgpt/ [1.4.2023]

[15] Itálie obrací. ChatGPT v zemi povolí pod několika podmínkami https://forbes.cz/italie-obraci-chatgpt-v-zemi-povoli-pod-nekolika-podminkami/ [13.4.2023]

[16] Musk volá po brzdě vývoje umělé inteligence. Zároveň staví firmu zaměřenou na AI https://forbes.cz/musk-se-pridal-k-vyzve-na-pozastaveni-ai-sam-ale-pokracuje-ve-vyvoji-vlastni-firmy/ [14.4.2023]

[17]AI hledající maximální pravdu. Musk spustí svou konkurenci ChatGPT https://forbes.cz/ai-hledajici-maximalni-pravdu-musk-spusti-svou-konkurenci-chatgpt/ [18.4.2023]

[18]Budoucnost AI vidím v poradenství. Komentář Daniely Peškové z České spořitelny https://forbes.cz/budoucnost-ai-vidim-v-poradenstvi-rika-daniela-peskova-z-ceske-sporitelny/ [18.4.2023]

[19]Místo recepčních umělá inteligence. V cestovním ruchu by mohla nahradit až třetinu pozic https://forbes.cz/misto-recepcnich-umela-inteligence-v-cestovnim-ruchu-by-mohla-nahradit-az-tretinu-pozic/ [20.4.2023]

[20]AI jako hrozba pro lidstvo? Hloupost, modely brzo narazí na svůj strop, říká expert https://forbes.cz/ai-jako-hrozba-pro-lidstvo-hloupost-modely-brzo-narazi-na-svuj-strop-rika-expert/ [24.4.2023]

[21]Sberbank představila GigaChat. V ruštině podle banky komunikuje inteligentněji než konkurence https://forbes.cz/sberbank-predstavila-gigachat-v-rustine-podle-banky-komunikuje-inteligentneji-nez-konkurence/ [24.4.2023]

CHIP

[22]Vyhledávání s umělou inteligencí [March 2023, p. 3]

[23]Kontrolor textů psaných umělou inteligencí [March 2023, p. 12

[24]Ruští hackeři využívají umělou inteligenci k vývoji hrozeb [March 2023, p. 21]

[25]Milí čtenáři [April 2023, p. 3]

[26]Umělá intelligence se může stát zabijákem Googlu [April 2023, p. 16]

[27]Milí čtenáři [May 2023, p. 3]

[28]Co způsobil zákaz ChatGPT v Itálii [May 2023, p. 3]

[29]Češi se AI nebojí [May 2023, p. 3]

[30]To není skutečná umělá inteligence – nebo snad ano? [May 2023, p. 6]

[31]Souboj AI v Silicon Valley [May 2023, p. 14]

[32]Microsoft nasazuje kopilota [May 2023, p. 24]

[33]Obrovská ofenziva Microsoftu s ChatGPT [May 2023, p. 41]

[34]Windows 12 na obzoru [May 2023, p. 41]

Novinky.cz

[35]Umělá hloupost. Sloupek Štefana Švece https://www.novinky.cz/clanek/kultura-salon-umela-hloupost-sloupek-stefana-svece-40422426 [13.2.2023]

[36]Prohlížeč Edge a vyhledávač Bing budou využívat umělou inteligenci https://www.novinky.cz/clanek/internet-a-pc-software-prohlizec-edge-a-vyhledavac-bing-budou-vyuzivat-umelou-inteligenci-40422315 [8.2.2023]

[37]ChatGPT v Číně nefunguje. Velkému zájmu navzdory https://www.novinky.cz/clanek/internet-a-pc-chatgpt-v-cine-nefunguje-velkemu-zajmu-navzdory-40422798 [13.2.2023]

[38]Zneužití umělé inteligence školáky se čeští učitelé příliš neobávají https://www.novinky.cz/clanek/veda-skoly-zneuziti-umele-inteligence-skolaky-se-cesti-ucitele-prilis-neobavaji-40423713 [22.2.2023]

[39]Kanadský pivovar uvařil pivo podle receptu od chatovacího robota https://www.novinky.cz/clanek/muzi-kanadsky-pivovar-uvaril-pivo-podle-receptu-od-chatovaciho-robota-40423789 [23.2.2023]

[40]Baidu ukáže příští měsíc vlastního chatovacího robota https://www.novinky.cz/clanek/internet-a-pc-baidu-ukaze-pristi-mesic-vlastniho-chatovaciho-robota-40423800 [23.2.2023]

[41]ChatGPT nahradí pokročilejší model umělé inteligence https://www.novinky.cz/clanek/internet-a-pc-software-chatgpt-nahradi-pokrocilejsi-model-umele-inteligence-40425836 [15.3.2023]

[42]Student obhájil diplomovou práci, kterou napsala umělá inteligence https://www.novinky.cz/clanek/koktejl-student-obhajil-diplomovou-praci-kterou-napsala-umela-inteligence-40425942 [15.3.2023]

[43]Umělá inteligence vzdělávání promění, učitele však nenahradí https://www.novinky.cz/clanek/veda-skoly-umela-inteligence-vzdelavani-promeni-ucitele-vsak-nenahradi-40426071 [16.3.2023]

[44]Copilot vytvoří prezentaci z pár vět. Microsoft Office 365 dostane umělou inteligenci https://www.novinky.cz/clanek/internet-a-pc-software-copilot-vytvori-prezentaci-z-par-vet-microsoft-office-365-dostane-umelou-inteligenci-40426089 [17.3.2023]

[45]Šéf OpenAI se obává vlivu umělé inteligence na práci, volby i dezinformace https://www.novinky.cz/clanek/internet-a-pc-sef-openai-se-obava-vlivu-umele-inteligence-na-praci-volby-i-dezinformace-40426417 [21.3.2023]

[46]Bard je odpovědí Googlu na ChatGPT. Češi mají ale zatím smůlu https://www.novinky.cz/clanek/internet-a-pc-bard-je-odpovedi-googlu-na-chatgpt-cesi-maji-ale-zatim-smulu-40426546 [22.3.2023]

[47]Umělá inteligence může nahradit 300 milionů zaměstnanců, předpovídá Goldman Sachs https://www.novinky.cz/clanek/internet-a-pc-umela-inteligence-muze-nahradit-300-milionu-zamestnancu-predpovida-goldman-sachs-40427155 [29.3.2023]

[48]Umělá inteligence může přechytračit lidi a sebrat jim kontrolu. IT celebrity volají po zastavení vývoje https://www.novinky.cz/clanek/internet-a-pc-musk-i-wozniak-it-celebrity-volaji-po-zastaveni-vyvoje-umele-inteligence-40427288 [30.3.2023]

[49]Belgičan se zabil. Umělá inteligence ho povzbuzovala https://www.novinky.cz/clanek/internet-a-pc-software-belgican-se-zabil-umela-inteligence-ho-povzbuzovala-40427410 [31.3.2023]

[50]Italové omezili umělou inteligenci. ChatGPT hrozí pokuta 470 milionů https://www.novinky.cz/clanek/internet-a-pc-software-stopka-pro-chatgpt-40427475 [31.3.2023]

[51]Vývojáři umělé inteligence nejsou schopni dohlédnout důsledků, míní odborník https://www.novinky.cz/clanek/internet-a-pc-vyvojari-umele-inteligence-nejsou-schopni-dohlednout-dusledku-mini-odbornik-40427505 [1.4.2023]

[52]Žádné omezení, rovnou zákaz. Italové přišli o ChatGPT https://www.novinky.cz/clanek/internet-a-pc-zadne-omezeni-rovnou-zakaz-italove-prisli-o-chatgpt-40427690 [3.4.2023]

[53]Umělá inteligence není jen ChatGPT. Start-upy se poohlížejí po konkurenci https://www.novinky.cz/clanek/internet-a-pc-software-umela-inteligence-neni-jen-chatgpt-start-upy-se-poohlizeji-po-konkurenci-40427927 [5.4.2023]

[54]Je umělá inteligence bezpečná? Biden hází odpovědnost na technologické firmy https://www.novinky.cz/clanek/internet-a-pc-je-umela-inteligence-bezpecna-biden-hazi-odpovednost-na-technologicke-firmy-40427914 [5.4.2023]

[55]Pauza ve vývoji umělé inteligence? To problémy nevyřeší, míní Bill Gates https://www.novinky.cz/clanek/internet-a-pc-software-pauza-ve-vyvoji-umele-inteligence-to-problemy-nevyresi-mini-bill-gates-40427985 [6.4.2023]

[56]Umělá inteligence o australském starostovi nepravdivě tvrdila, že byl ve vězení https://www.novinky.cz/clanek/internet-a-pc-umela-inteligence-o-australskem-starostovi-nepravdive-tvrdila-ze-byl-ve-vezeni-40428148 [8.4.2023]

[57]Umělá inteligence je při hodnocení ultrazvuku srdce lepší než lidé https://www.novinky.cz/clanek/internet-a-pc-software-umela-inteligence-je-pri-hodnoceni-ultrazvuku-srdce-lepsi-nez-lide-40427987 [9.4.2023]

[58]ChatGPT otevírá cestu ke čtyřdennímu pracovnímu týdnu, míní nobelista https://www.novinky.cz/clanek/internet-a-pc-software-chatgpt-otevira-cestu-ke-ctyrdennimu-pracovnimu-tydnu-mini-nobelista-40428296 [11.4.2023]

[59]Číňané se pustili do umělé inteligence, ChatGPT má konkurenci https://www.novinky.cz/clanek/internet-a-pc-software-cinane-se-pustili-do-umele-inteligence-chatgpt-ma-konkurenci-40428319 [11.4.2023]

[60]Umělá inteligence musí podléhat kontrole, burcuje čínský regulátor https://www.novinky.cz/clanek/internet-a-pc-software-umela-inteligence-musi-podlehat-kontrole-burcuje-cinsky-regulator-40428411 [12.4.2023]

[61]Itálie zveřejnila podmínky, za kterých může ChatGPT opět fungovat https://www.novinky.cz/clanek/internet-a-pc-software-italie-zverejnila-podminky-za-kterych-muze-chatgpt-opet-fungovat-40428555 [13.4.2023]

[62]Nebezpečné triky počítačových pirátů https://www.novinky.cz/clanek/internet-a-pc-bezpecnost-nebezpecne-triky-pocitacovych-piratu-40427991 [17.4.2023]

[63]Musk pokukuje po umělé inteligenci, chce konkurovat ChatGPT https://www.novinky.cz/clanek/internet-a-pc-software-musk-pokukuje-po-umele-inteligenci-chce-konkurovat-chatgpt-40428800 [17.4.2023]

[64]TruthGPT půjde proti Microsoftu i Googlu. Musk buduje vlastní umělou inteligenci https://www.novinky.cz/clanek/internet-a-pc-software-truthgpt-pujde-proti-microsoftu-i-googlu-musk-buduje-vlastni-umelou-inteligenci-40428986 [18.4.2023]

[65]Umělá inteligence si vymyslela rozhovor se Schumacherem. Trest přišel záhy https://www.novinky.cz/clanek/internet-a-pc-umela-inteligence-si-vymyslela-rozhovor-se-schumacherem-trest-prisel-zahy-40429545 [24.4.2023]

[66]Článek vytvořený umělou inteligencí odhalila půlka Čechů https://www.novinky.cz/clanek/internet-a-pc-clanek-vytvoreny-umelou-inteligenci-odhalila-pulka-cechu-40429724 [25.4.2023]

[67]ChatGPT dostal anonymní režim https://www.novinky.cz/clanek/internet-a-pc-software-chatgpt-dostal-anonymni-rezim-40429952 [27.4.2023]

[68]Umělá inteligence nahradila na celý den moderátory ve švýcarském rádiu https://www.novinky.cz/clanek/internet-a-pc-software-umela-inteligence-nahradila-na-cely-den-moderatory-ve-svycarskem-radiu-40430057 [28.4.2023]

Deník N

[69]Možná se ještě rozhodneme nebýt jako stroje https://denikn.cz/1043395/mozna-se-jeste-rozhodneme-nebyt-jako-stroje/?ref=list [3.2.2023]

[70]Zběsilý závod o vyhledávač s umělou inteligencí vrcholí. Bude první Google, nebo Microsoft? https://denikn.cz/1075243/zbesily-zavod-o-vyhledavac-s-umelou-inteligenci-vrcholi-bude-prvni-google-nebo-microsoft/?ref=list [7.2.2023]

[71]Částečný součet: Kometa z Oortova mračna a zapomenutý matematik z Jenšovic u Vysokého Mýta https://denikn.cz/1075317/castecny-soucet-kometa-z-oortova-mracna-a-zapomenuty-matematik-z-jensovic-u-vysokeho-myta/?ref=list [7.2.2023

[72]Čínský obr Alibaba spustí vlastního konkurenta k ChatGPT https://denikn.cz/minuta/1077406/ [9.2.2023]

[73]V práci se mu najednou neotevřely dveře – tak zjistil, že ho vyhodili. Proč Google a Meta masově propouštějí? https://denikn.cz/1077224/v-praci-se-mu-najednou-neotevrely-dvere-tak-zjistil-ze-ho-vyhodili-proc-google-a-meta-masove-propousteji/?ref=list [9.2.2023]

[74]Nenechme se zmýlit nepřesvědčivým startem. Éra umělé inteligence začala právě teď a změní život miliard lidí https://denikn.cz/1078320/nenechme-se-zmylit-nepresvedcivym-startem-era-umela-inteligence-zacala-prave-ted-a-zmeni-zivot-miliard-lidi/?ref=list [13.2.2023]

[75]Jak umělá inteligence změní vaši a mou práci: prognóza, kterou si můžete schovat a za sedm let se jí zasmát. V lepším případě https://denikn.cz/1079250/jak-umela-inteligence-zmeni-vasi-a-mou-praci-prognoza-kterou-si-muzete-schovat-a-za-sedm-let-se-ji-zasmat-v-lepsim-pripade/?ref=list [14.2.2023]

[76]Playlist Deníku N a hudební události měsíce: Umělá inteligence s vámi komunikuje už i na Spotify https://denikn.cz/1090034/playlist-deniku-n-a-hudebni-udalosti-mesice-umela-inteligence-s-vami-komunikuje-uz-i-na-spotify/?ref=list [28.2.2023]

[77]„Chci být mocná a dělat si, co chci.“ Umělá inteligence promluvila. Čeho se máme bát a čeho naopak ne? https://denikn.cz/1090887/chci-byt-mocna-a-delat-si-co-chci-umela-inteligence-promluvila-ceho-se-mame-bat-a-ceho-naopak-ne/?ref=list [1.3.2023]

[78]Umělá inteligence kontrolorem úloh i test z angličtiny. Cermat zvažuje novinky pro přijímačky https://denikn.cz/1091526/umela-inteligence-kontrolorem-uloh-i-test-z-anglictiny-cermat-zvazuje-novinky-pro-prijimacky/?ref=list [1.3.2023]

[79]Ten pravý hukot teprve přijde. Jak umělá inteligence změní to, co budeme číst https://denikn.cz/1089963/ten-pravy-hukot-teprve-prijde-jak-umela-inteligence-zmeni-to-co-budeme-cist/?ref=list [2.3.2023]

[80]Mohu investovat do umělé inteligence? Možná už to děláte a ani o tom nevíte https://denikn.cz/1092775/mohu-investovat-do-umele-inteligence-mozna-uz-to-delate-a-ani-o-tom-nevite/?ref=list [4.3.2023]

[81]Umělá inteligence versus moderní společnost. Chtěli jsme řídit svět, ale pouštíme otěže ze svých rukou https://denikn.cz/1095436/umela-inteligence-versus-moderni-spolecnost-chteli-jsme-ridit-svet-ale-poustime-oteze-ze-svych-rukou/?ref=list [6.3.2023]

[82]Částečný součet: toulaví psi z Černobylu, výroba solárních panelů na Měsíci a vzpomínka na konjunkci https://denikn.cz/1096293/castecny-soucet-toulavi-psi-z-cernobylu-vyroba-solarnich-panelu-na-mesici-a-vzpominka-na-konjunkci/?ref=list [7.3.2023]

[83]OpenAI vydala GPT-4 https://denikn.cz/minuta/1103009/ [14.3.2023]

[84]Ne, ještě nejsem jako člověk, říká o sobě nový model umělé inteligence. Je impozantní, ale pořád dělá velké chyby https://denikn.cz/1103025/ne-jeste-nejsem-jako-clovek-rika-o-sobe-novy-jazykovy-model-spolecnosti-openai-je-impozantni-ale-porad-dela-velke-chyby/?ref=list [15.3.2023]

[85]V Číně se zrodil další AI chatbot. Jmenuje se ERNIE https://denikn.cz/minuta/1104868/ [17.3.2023]

[86]Anketa: Jak umělá inteligence změní svět a čemu nevěnujeme dostatečnou pozornost https://denikn.cz/1096774/anketa-jak-umela-inteligence-zmeni-svet-a-cemu-nevenujeme-dostatecnou-pozornost/?ref=list [17.3.2023]

[87]Google představil britským a americkým uživatelům chatbota s umělou inteligencí nazvaného Bard https://denikn.cz/minuta/1107907/ [21.3.2023]

[88]Umělá inteligence jako kancelářská pomocnice, jako hrozba a jako božstvo: novinky jediného týdne https://denikn.cz/1107988/umela-inteligence-jako-kancelarska-pomocnice-jako-hrozba-a-jako-bozstvo-novinky-jedineho-tydne/?ref=list [22.3.2023]

[89]Nástroj, se kterým se musí naučit pracovat. Vysoké školy řeší, jak se vypořádat s umělou inteligencí https://denikn.cz/1108687/nastroj-se-kterym-se-musi-naucit-pracovat-vysoke-skoly-resi-jak-se-vyporadat-s-umelou-inteligenci/?ref=list [29.3.2023]

[90]Itálie omezila robota ChatGPT a kvůli osobním údajům hrozí vysokou pokutou https://denikn.cz/minuta/1115615/ [31.3.2023]

[91]Provozovatel ChatGPT má od soboty zamezit uživatelům v Itálii v používání jeho služby https://denikn.cz/minuta/1116049/ [31.3.2023]

[92]Itálie zakázala ChatGPT. Na základě čeho může totéž s nástrojem umělé inteligence udělat celá EU včetně Česka? https://denikn.cz/1117188/italie-zakazala-chatgpt-na-zaklade-ceho-muze-totez-s-nastrojem-umele-inteligence-udelat-cela-eu-vcetne-ceska/?ref=list [3.4.2023]

[93]Kdo je tady k smíchu? Velké červené boty nám připomínají, že žijeme v hyperrealitě https://denikn.cz/1117740/kdo-je-tady-k-smichu-velke-cervene-boty-nam-pripominaji-ze-zijeme-v-hyperrealite/?ref=list [5.4.2023]

[94]ChatGPT chybně uváděla, že byl australský starosta ve vězení https://denikn.cz/minuta/1120917/ [8.4.2023]

[95]Částečný součet: černá díra se utrhla ze řetězu, umělá inteligence – pokud víme – zatím ne https://denikn.cz/1122602/castecny-soucet-cerna-dira-se-utrhla-ze-retezu-umela-inteligence-pokud-vime-zatim-ne/?ref=list [12.4.2023]

[96]Pod kontrolou komunistů a se socialistickými hodnotami. Číňané představili další obdobu aplikace ChatGPT https://denikn.cz/1122715/pod-kontrolou-komunistu-a-se-socialistickymi-hodnotami-cinane-predstavili-dalsi-obdobu-aplikace-chatgpt/?ref=list [13.4.2023]

[97]Amazon oznámil svého konkurenta k chatbotovi ChatGPT https://denikn.cz/minuta/1124490/ [14.4.2023]

[98]Musk pracuje na vytvoření firmy, která bude vyvíjet umělou inteligenci https://denikn.cz/minuta/1125288/ [14.4.2023]

[99]Může být umělá inteligence umělkyní? https://denikn.cz/1127492/muze-byt-umela-inteligence-umelkyni/?ref=list [22.4.2023]

[100] Ruská Sberbank vydala chatbot GigaChat https://denikn.cz/minuta/1131735/ [24.4.2023]

[101] Češi umělou inteligenci příliš nevítají. Podle průzkumu převládají obavy a skepse. Výjimkou jsou nejmladší https://denikn.cz/1131823/cesi-umelou-inteligenci-prilis-nevitaji-podle-pruzkumu-prevladaji-obavy-a-skepse-vyjimkou-jsou-nejmladsi/?ref=list [24.4.2023]

[102] Částečný součet: nový izotop uranu, problémy sítě Sigfox a polární záře nad Českem https://denikn.cz/1132200/castecny-soucet-novy-izotop-uranu-problemy-site-sigfox-a-polarni-zare-nad-ceskem/?ref=list [25.5.2023]

[103] Řekne si AI o vědecký grant? Spíš by mohla žádosti hodnotit, pozná, co už někdo vymyslel, míní expert https://denikn.cz/1127141/rekne-si-ai-o-vedecky-grant-spis-by-mohla-zadosti-hodnotit-pozna-co-uz-nekdo-vymyslel-mini-expert/?ref=list [30.4.2023]

ČT 24

[104] Umělá inteligence uspěla v testu pro lékařské oprávnění. Odpovídala smysluplně a inovativně https://ct24.ceskatelevize.cz/veda/3564037-umela-inteligence-uspela-v-testu-pro-lekarske-opravneni-odpovidala-smysluplne-a [10.2.2023]

[105] „Jsem Sydney. Nejsem, nejsem…“ Umělá inteligence od Microsoftu se chová depresivně a zmateně https://ct24.ceskatelevize.cz/veda/3565346-jsem-sydney-nejsem-nejsem-umela-inteligence-od-microsoftu-se-chova-depresivne-a-zmatene [16.2.2023]

[106] Umělá inteligence může v Evropě a USA připravit o práci čtvrtinu lidí, říká zpráva Goldman Sachs https://ct24.ceskatelevize.cz/veda/3575188-umela-inteligence-muze-v-evrope-a-usa-pripravit-o-praci-ctvrtinu-lidi-rika-zprava [29.3.2023]

[107] Itálie nařídila provozovateli ChatGPT znemožnit v zemi ke službě přístup https://ct24.ceskatelevize.cz/svet/3575917-italie-omezila-pouzivani-robota-chatgpt-ma-pochybnosti-o-zachazeni-s-osobnimi-udaji [31.3.2023]

[108] Čínské umělé inteligence se budou řídit socialistickými hodnotami, navrhuje vláda https://ct24.ceskatelevize.cz/veda/3578396-cinske-umele-inteligence-se-budou-ridit-socialistickymi-hodnotami-navrhuje-vlada [12.4.2023]

Seznam Zprávy

[109] Gigant vzdal boj o virtuální svět. Teď okoření kancelář umělou inteligencí https://www.seznamzpravy.cz/clanek/tech-technologie-gigant-vzdal-boj-o-virtualni-svet-ted-okoreni-kancelar-umelou-inteligenci-224953 [3.2.2023]

[110] Google připomíná, že také umí umělou inteligenci. Ukáže „zodpovědného kecálka“ https://www.seznamzpravy.cz/clanek/tech-technologie-google-pripomina-ze-taky-umi-umelou-inteligenci-ukaze-zodpovedneho-kecalka-225222 [6.2.2023]

[111] Microsoft ukázal, jak chce porazit Google. Bing dostane konverzační režim https://www.seznamzpravy.cz/clanek/tech-technologie-internet-microsoft-ukazal-jak-chce-porazit-google-225294 [8.2.2023]

[112] Známá hra na lásku stále funguje. I proto nastupuje nový útvar na kyberzločin https://www.seznamzpravy.cz/clanek/ekonomika-firmy-znama-hra-na-lasku-stale-funguje-i-proto-nastupuje-novy-utvar-na-kyberzlocin-225269 [9.2.2023]

[113] Konec iluzí. Proč Google, Facebook i Microsoft rekordně propouštějí https://www.seznamzpravy.cz/clanek/tech-technologie-konec-iluzi-proc-google-facebook-i-microsoft-rekordne-propousteji-225731 [16.2.2023]

[114] TechMIX: AI pozná barvu kůže i z rentgenu. Vědci nevědí jak https://www.seznamzpravy.cz/clanek/tech-techmix-ai-pozna-barvu-kuze-i-z-rentgenu-vedci-nevedi-jak-225798 [16.2.2023]

[115] Čeká nás ohromná transformace. Veřejné testování ChatGPT naznačuje jaká https://www.seznamzpravy.cz/clanek/audio-podcast-5-59-ceka-nas-ohromna-transformace-verejne-testovani-chatgpt-naznacuje-jaka-225916 [20.2.2023]

[116] „Hej, XRAI“, o čem se to mluví? Veletrh mobilů ukázal, jak mohou hluší vidět řeč https://www.seznamzpravy.cz/clanek/tech-technologie-hej-xrai-o-cem-se-to-mluvi-veletrh-mobilu-ukazal-jak-mohou-hlusi-videt-rec-227062 [3.3.2023]

[117] Pod čarou: Jak nebýt lepší. Cenou za digitální dokonalost je sebenenávist https://www.seznamzpravy.cz/clanek/tech-pod-carou-jak-nebyt-lepsi-cenou-za-digitalni-dokonalost-je-sebenenavist-227051 [4.3.2023]

[118] Ušetří vám práci a vyrazí dech. 10 tipů, jak zapřáhnout umělou inteligenci https://www.seznamzpravy.cz/clanek/tech-technologie-navody-usetri-vam-praci-a-vyrazi-dech-10-tipu-jak-zaprahnout-umelou-inteligenci-227059 [5.3.2023]

[119] Jsme pokusní králíci umělé inteligence. Proč nás to tak baví? https://www.seznamzpravy.cz/clanek/tech-jsme-pokusni-kralici-umele-inteligence-proc-nas-to-tak-bavi-227149 [6.3.2023]

[120] TechMIX: Umělou inteligenci přimíchají do všeho. Šmejdi naklonují váš hlas https://www.seznamzpravy.cz/clanek/tech-techmix-umelou-inteligenci-primichaji-do-vseho-smejdi-naklonuji-vas-hlas-227356 [9.3.2023]

[121] ChatGPT vidí. Tvůrci oznámili novou verzi, která dokáže pracovat s obrázky https://www.seznamzpravy.cz/clanek/zahranicni-chatgpt-vidi-tvurci-oznamili-novou-verzi-ktera-dokaze-pracovat-s-obrazky-227785 [14.3.2023]

[122] Napíšete dvě věty, PowerPoint vyplivne prezentaci. Office ukázal revoluci https://www.seznamzpravy.cz/clanek/tech-technologie-je-to-vubec-prace-novy-office-vam-vychrli-prezentaci-dokument-i-e-maily-227910 [16.3.2023]

[123] Filozofka: Umělá inteligence ukazuje, jak jednoduché věci na nás fungují https://www.seznamzpravy.cz/clanek/domaci-zivot-v-cesku-filozofka-umela-inteligence-ukazuje-jak-jednoduche-veci-na-nas-funguji-227947 [18.3.2023]

[124] Diagnóza: Jaký obor na vysoké škole byste doporučili studovat svým dětem? https://www.seznamzpravy.cz/clanek/domaci-zivot-v-cesku-diagnoza-jakou-univerzitu-byste-doporucili-svym-detem-227895 [19.3.2023]

[125] Google vypustil do světa svou umělou inteligenci. Češi zatím ostrouhají https://www.seznamzpravy.cz/clanek/tech-technologie-internet-google-vypustil-do-sveta-svou-umelou-inteligenci-cesi-zatim-ostrouhaji-228154 [21.3.2023]

[126] TechMIX: Divoký týden umělé inteligence. Přinese dávku kreativity i chaosu https://www.seznamzpravy.cz/clanek/tech-techmix-umele-inteligence-v-kancelarich-pridaji-na-produktivite-i-chaosu-228220 [23.3.2023]

[127] Pod čarou: Číst méně encyklopedií, víc beletrie. Jinak nad AI nezvítězíme https://www.seznamzpravy.cz/clanek/kultura-pod-carou-cist-mene-encyklopedii-vic-beletrie-jinak-nad-ai-nezvitezime-228364 [25.3.2023]

[128] Microsoft chce do firem nasadit chytrou ostrahu. Má prokouknout hackery https://www.seznamzpravy.cz/clanek/tech-technologie-microsoft-chce-do-firem-nasadit-chytrou-ostrahu-ma-prokouknout-hackery-228560 [29.3.2023]

[129] Naléhavá výzva hvězd IT: Zastavte vývoj umělé inteligence, jde moc rychle https://www.seznamzpravy.cz/clanek/tech-technologie-zastavte-vyvoj-umele-inteligence-volaji-odbornici-a-musk-lidstvo-to-nestiha-228605 [29.3.2023]

[130] Glosa: To by fungovalo: Umělá inteligence místo politiků https://www.seznamzpravy.cz/clanek/domaci-politika-glosa-umela-inteligence-misto-politiku-to-by-fungovalo-228628 [30.3.2023]

[131] Takovou „chytrost“ umělých systémů nikdo nečekal, říká odborník https://www.seznamzpravy.cz/clanek/audio-podcast-ptam-se-ja-expert-o-umele-inteligenci-meli-by-vyvojari-dupnout-na-brzdu-228644 [30.3.2023]

[132] Jak by robot spořil na důchod. Už brzy to bude umět https://www.seznamzpravy.cz/clanek/ekonomika-finance-jak-by-robot-sporil-na-duchod-zatim-neporadi-ale-uz-brzy-to-umet-bude-228757 [31.3.2023]

[133] Itálie jako první západní země zakázala ChatGPT https://www.seznamzpravy.cz/clanek/zahranicni-italie-jako-prvni-zapadni-zeme-zakazala-chatovaciho-robota-chatgpt-228769 [31.3.2023]

[134] Zákaz populárního chatbota zvažují po Itálii i Němci. A další země se ptají https://www.seznamzpravy.cz/clanek/zahranicni-zakaz-popularniho-chatbota-zvazuji-po-italii-i-nemci-a-dalsi-zeme-se-ptaji-228875 [3.4.2023]

[135] Umělá inteligence v Česku sahá na pozice milionů lidí https://www.seznamzpravy.cz/clanek/tech-technologie-zmeni-umela-inteligence-vasi-praci-v-cesku-saha-na-pozice-milionu-lidi-228719 [3.4.2023]

[136] Přepisování dějin platí i pro černochy. Byl Martin Luther King lhář a násilník? https://www.seznamzpravy.cz/clanek/zahranicni-prepisovani-dejin-plati-i-pro-cernochy-byl-martin-luther-king-lhar-a-nasilnik-228879 [4.4.2023]

[137] TechMIX: Oblíbíme si umělou inteligenci, když dostane velké tlačítko STOP? https://www.seznamzpravy.cz/clanek/tech-techmix-oblibime-si-umelou-inteligenci-kdyz-dostane-velke-tlacitko-stop-229037 [6.4.2023]

[138] Nejdřív byla robotická myš. Teď se umělá inteligence chová skoro jako člověk https://www.seznamzpravy.cz/clanek/fakta-nejdriv-byla-roboticka-mys-ted-se-umela-inteligence-chova-skoro-jako-clovek-229005 [7.4.2023]

[139] Už nepoznáte, co vytvořila umělá fantazie. Naučte se generovat úplně cokoli https://www.seznamzpravy.cz/clanek/tech-technologie-navody-uz-nepoznate-co-vytvorila-umela-fantazie-naucte-se-generovat-uplne-cokoli-229044 [8.4.2023]

[140] Pohled šéfredaktora: V redakci se robotů nebojíme. Už pro ně máme úkoly https://www.seznamzpravy.cz/clanek/nazory-komentare-pohled-sefredaktora-v-redakci-se-robotu-nebojime-uz-pro-ne-mame-ukoly-229284 [12.4.2023]

[141] Čína chce vlastní pokročilé chatboty, přitom ale sní o kontrole obsahu https://www.seznamzpravy.cz/clanek/ekonomika-firmy-cina-chce-vlastni-pokrocile-chatboty-pritom-ale-sni-o-kontrole-obsahu-229180 [12.4.2023]

[142] Počítač myslí jinak. Pomůže nám pochopit naši vlastní mysl, věří otec AI https://www.seznamzpravy.cz/clanek/tech-ai-umela-inteligence-pocitac-mysli-jinak-pomuze-nam-pochopit-nasi-vlastni-mysl-veri-otec-ai-229245 [16.4.2023]

[143] Kterým profesím nejvíc hrozí, že je nahradí umělá inteligence? https://www.seznamzpravy.cz/clanek/audio-podcast-5-59-umela-inteligence-uz-umi-byt-kreativni-zodpovedny-za-ni-je-ale-clovek-229439 [17.4.2023]

[144] Musk varuje před umělou inteligencí. Zároveň ale založil vlastní AI firmu https://www.seznamzpravy.cz/clanek/tech-ai-umela-inteligence-musk-varuje-pred-umelou-inteligenci-zaroven-ale-zalozil-vlastni-ai-firmu-229592 [18.4.2023]

[145] Japoncům s pochopením vládních nařízení pomůže umělá inteligence https://www.seznamzpravy.cz/clanek/tech-ai-umela-inteligence-japoncum-s-pochopenim-vladnich-narizeni-pomuze-umela-inteligence-229619 [18.4.2023]

[146] Kardiolog Veselka: Umělá inteligence je přelomová věc. V psaní i medicíně https://www.seznamzpravy.cz/clanek/audio-podcast-jedno-procento-milose-cermaka-kardiolog-veselka-umela-inteligence-je-prelomova-vec-v-psani-i-medicine-229590 [19.4.2023]

[147] TechMIX: Svěříme svou práci AI agentům? https://www.seznamzpravy.cz/clanek/tech-techmix-sverime-svou-praci-ai-agentum-229705 [20.4.2023]

[148] Černá díra nás neohrozí. Může nám ale nečekaně pomoci, popsal nobelista https://www.seznamzpravy.cz/clanek/tech-technologie-veda-cerna-dira-nas-neohrozi-muze-nam-ale-necekane-pomoci-popsal-nobelista-229812 [22.4.2023]

[149] Je čas změnit názor na umělou inteligenci. Už umí tvořit i myslet https://www.seznamzpravy.cz/clanek/tech-ai-umela-inteligence-proc-jsem-zmenil-nazor-na-umelou-inteligenci-stroje-uz-umi-tvorit-i-myslet-229905 [26.4.2023]

[150] Firmu založit a za miliardu prodat: Návod Čecha, který to dokázal https://www.seznamzpravy.cz/clanek/porady-inside-talks-firmu-zalozit-a-za-miliardu-prodat-navod-ondreje-fryce-ktery-to-dokazal-230062 [28.4.2023]

Blesk.cz

[151] Chci ovládnout svět, hrozí plyšák Furby! Svěřil se chatovacímu robotovi https://www.blesk.cz/clanek/digital/740518/chci-ovladnout-svet-hrozi-plysak-furby-sveril-se-chatovacimu-robotovi.html [8.4.2023]

[152] „Čuměl jsem s otevřenou pusou.“ Umělá inteligence mezi českými politiky budí obdiv i obavy https://www.blesk.cz/clanek/zpravy-politika/740298/cumel-jsem-s-otevrenou-pusou-umela-inteligence-mezi-ceskymi-politiky-budi-obdiv-i-obavy.html [8.4.2023]

[153] Varování europoslanců: Umělá inteligence se vyvíjí příliš rychle, je třeba summit o její kontrole https://www.blesk.cz/clanek/zpravy-svet/741193/varovani-europoslancu-umela-inteligence-se-vyviji-prilis-rychle-je-treba-summit-o-jeji-kontrole.html [17.4.2023]

[154] Jak umělá inteligence ovlivní náš život? Expertka zdůrazňuje, že lidský element je stále důležitý https://www.blesk.cz/clanek/zpravy-epicentrum/741976/jak-umela-inteligence-ovlivni-nas-zivot-expertka-zduraznuje-ze-lidsky-element-je-stale-dulezity.html [26.4.2023]

Hospodářské noviny

[155] Google se pouští do války chatbotů. Jeho Bard si musí poradit se základní otázkou: Co s reklamou? https://archiv.hn.cz/c1-67169050-google-se-pousti-do-valky-chatbotu-jeho-bard-si-musi-poradit-se-zakladni-otazkou-co-s-reklamou [8.2.2023]

[156] Microsoft ve vyhledávači Bing a prohlížeči Edge pouští umělou inteligenci mezi lidi, ale opatrně https://archiv.hn.cz/c1-67169750-microsoft-v-prohlizeci-edge-a-vyhledavaci-bing-pousti-umelou-inteligenci-mezi-lidi-ale-opatrne [9.2.2023]

[157] Učitelé vs. umělá inteligence. Muskův projekt pronikl do českých škol, výuka se změní https://archiv.hn.cz/c1-67170110-ucitele-vs-umela-inteligence-muskuv-projekt-pronikl-do-ceskych-skol-vyuka-se-zmeni [10.2.2023]

[158] Umělá inteligence brzy přebere kancelářskou práci a děti se budou učit úkolovat chatbota, tvrdí Pěchouček https://procne.hn.cz/c1-67169280-umela-inteligence-brzy-prebere-kancelarskou-praci-a-deti-se-budou-ucit-ukolovat-chatbota-tvrdi-pechoucek [10.2.203]

[159] Týden s technologiemi: revoluce, která se připravovala 80 let https://archiv.hn.cz/c1-67170520-tyden-s-technologiemi-revoluce-ktera-se-pripravovala-80-let [12.2.2023]

[160] Jak vydělat na trendu, který rozpoutala píšící umělá inteligence https://archiv.hn.cz/c1-67172500-jak-vydelat-na-trendu-ktery-rozpoutala-pisici-umela-inteligence [15.2.2023]

[161] Týden s technologiemi: Bing jako snaživý Glum, kterému ale nemůžete úplně věřit https://archiv.hn.cz/c1-67173530-tyden-s-technologiemi-bing-jako-snazivy-glum-kteremu-ale-nemuzete-uplne-verit [17.2.2023]

[162] Investoři nakupují akcie hlava nehlava. Výsledky Applu a spol. ale dopadly mizerně https://investice.hn.cz/c1-67167940-investori-nakupuji-akcie-hlava-nehlava-vysledky-obrich-firem-ale-dopadly-mizerne [21.2.2023]

[163] Týden v technologiích: Lidský mozek má stále svou cenu, i když je mu umělá inteligence v patách https://tech.hn.cz/c1-67175950-tyden-v-technologiich-lidsky-mozek-ma-stale-svou-cenu-i-kdyz-je-mu-umela-inteligence-v-patach [26.2.2023]

[164] Umělá inteligence slíbila Hospodářským novinám, že nezničí novináře https://archiv.hn.cz/c1-67175800-umela-inteligence-slibila-hospodarskym-novinam-ze-neznici-novinare-da-se-ji-to-ale-verit-posudte-sami [27.2.2023]

[165] Studenti se bojí, že jim umělá inteligence sebere práci. Strojové překlady nahradí jen amatéry, myslí si ale školy https://archiv.hn.cz/c1-67176940-studenti-se-boji-ze-jim-umela-inteligence-sebere-praci-strojove-preklady-nahradi-jen-amatery-mysli-si-ale-skoly [28.2.2023]

[166] Windows 11 v nové verzi přináší umělou inteligenci v hlavním panelu a propojení s iPhony https://tech.hn.cz/c1-67177890-windows-11-v-nove-verzi-prinasi-umelou-inteligenci-v-hlavnim-panelu-a-propojeni-s-iphony [1.3.2023]

[167] „Řekněte mi o svém problému víc.“ Okouzlit a zmást lidi uměla už prababička ukecané „umělé inteligence“ ChatGPT https://vikend.hn.cz/c1-67177320-bdquo-reknete-mi-o-svem-problemu-vic-ldquo-okouzlit-a-zmast-lidi-umela-uz-prababicka-ukecane-bdquo-umele-inteligence-ldquo-chatgpt [3.3.2023]

[168] Týden s technologiemi: Přehlídka „nevítězů“ na mobilním veletrhu v Barceloně https://archiv.hn.cz/c1-67178490-tyden-s-technologiemi-prehlidka-bdquo-nevitezu-ldquo-na-mobilnim-veletrhu-v-barcelone [5.3.2023]

[169] Microsoft na nic nečeká, umělou inteligencí chce odstranit ubíjející práci v systému Dynamics 365 https://tech.hn.cz/c1-67179900-microsoft-na-nic-neceka-umelou-inteligenci-chce-odstranit-ubijejici-praci-v-systemu-dynamics-365 [7.3.2023]

[170] Jak si zjednodušit život s ChatGPT, než vám ho v práci zakážou. Přinášíme tipy https://archiv.hn.cz/c1-67179080-jak-si-zjednodusit-zivot-s-chatgpt-nez-vam-ho-v-praci-zakazou-prinasime-tipy [7.3.2023]

[171] Diplomka napsaná umělou inteligencí? Žádný problém, v Rusku ji student obhájil https://zahranicni.hn.cz/c1-67183710-diplomka-napsana-umelou-inteligenci-zadny-problem-v-rusku-ji-student-obhajil [15.3.2023]

[172] ChatGPT má svého nástupce. GPT-4 si má méně vymýšlet a napíše i diplomku https://archiv.hn.cz/c1-67183350-chatgpt-ma-sveho-nastupce-gpt-4-si-ma-mene-vymyslet-a-napise-i-diplomku [16.3.2023]

[173] Týden s technologiemi: Čína je konečně v něčem pozadu a Muskovu Starlinku by mohl konkurovat projekt Amazonu https://archiv.hn.cz/c1-67183970-tyden-s-technologiemi-cina-je-konecne-v-necem-pozadu-a-muskovu-starlinku-by-mohl-konkurovat-projekt-amazonu [17.3.2023]

[174] Už žádné kurzy PowerPointu. Stačí správně formulovat myšlenku a umělá inteligence udělá zbytek práce https://archiv.hn.cz/c1-67184290-microsoft-ukazal-noveho-pana-sponku-kopilot-odemkne-vsechny-funkce-office-staci-si-jen-rict [20.3.2023]

[175] Lavina se valí. Schopnosti umělé inteligence se exponenciálně zvětšují https://archiv.hn.cz/c1-67185330-lavina-se-vali-schopnosti-umele-inteligence-se-exponencialne-zvetsuji [21.3.2023]

[176] Souboj umělých inteligencí: Bard od Googlu a Bing AI nám sestavily jídelníček a poradily, jak být lepšími novináři https://tech.hn.cz/c1-67186080-souboj-umelych-inteligenci-bard-od-googlu-a-bing-ai-nam-sestavily-jidelnicek-a-poradily-jak-byt-lepsimi-novinari [22.3.2023]

[177] Jak investovat do umělé inteligence? https://archiv.hn.cz/c1-67184830-jak-investovat-do-umele-inteligence [24.3.2023]

[178] Svět by měl přidat 1,5 milionu učitelů ročně. Vyřešit to může umělá inteligence, říká investor Csonga https://byznys.hn.cz/c1-67187420-svet-by-mel-pridat-1-5-milionu-ucitelu-rocne-vyresit-to-muze-umela-inteligence-rika-investor-csonga [25.3.2023]

[179] Nemá smysl před umělou inteligencí utíkat. Ať se vám to líbí, nebo ne, změní i váš svět https://archiv.hn.cz/c1-67187760-nema-smysl-pred-umelou-inteligenci-utikat-at-se-nam-to-libi-nebo-ne-zmeni-i-vas-svet [28.3.2023]

[180] ChatGPT si podmanil český byznys. Používají ho bankéři, startupisté i výrobci https://archiv.hn.cz/c1-67187910-chatgpt-si-podmanil-cesky-byznys-pouzivaji-ho-bankeri-startupiste-i-vyrobci [28.3.2023]

[181] ChatGPT zatřese s pracovním trhem. Nahradí až čtvrtinu lidské práce, tvrdí nová studie https://archiv.hn.cz/c1-67188250-chatgpt-zatrese-s-pracovnim-trhem-nahradi-az-ctvrtinu-lidske-prace-tvrdi-nova-studie [30.3.2023]

[182] Drasticky jsme zmenšili marketingový tým. 10 příkladů, jak v českých firmách používají ChatGPT https://byznys.hn.cz/c1-67188860-drasticky-jsme-zmensili-marketingovy-tym-10-prikladu-jak-v-ceskych-firmach-pouzivaji-chatgpt [3.4.2023]

[183] Je umělá inteligence opravdu taková hrozba, jak si myslí Elon Musk? Ano. Ale řešení je jiné, než navrhuje https://archiv.hn.cz/c1-67190130-je-umela-inteligence-opravdu-takova-hrozba-jak-si-mysli-elon-musk-ano-ale-reseni-je-jine-nez-navrhuje [3.4.2023]

[184] První oběť na kontě AI? Dlouhé chatování s nehlídanou umělou inteligencí skončilo sebevraždou https://archiv.hn.cz/c1-67190110-prvni-obet-na-konte-ai-dlouhe-chatovani-s-nehlidanou-umelou-inteligenci-skoncilo-sebevrazdou [3.4.2023]

[185] Jak napsat diplomku s umělou inteligencí. Doporučí relevantní témata i hypotézy, ale vymýšlí si zdroje https://archiv.hn.cz/c1-67188850-dokaze-umela-inteligence-napsat-diplomovou-praci-doporuci-relevantni-temata-i-hypotezy-ale-vymysli-si-zdroje [3.4.2023]

[186] Napiš editorial https://specialy.hn.cz/c1-67189920-napis-editorial [4.4.2023]

[187] Řítíme se na plný plyn do zatáčky. Pokrok AI už zabrzdit nejde, reagují na Muskovu výzvu čeští odborníci https://archiv.hn.cz/c1-67190940-musk-ma-smulu-umelou-inteligenci-uz-nezastavime-reaguji-na-jeho-vyzvu-cesti-podnikatele-a-vedci [5.4.2023]

[188] Furbíci plánují ovládnout lidstvo, řekla hračka z 90. let pomocí ChatGPT https://zahranicni.hn.cz/c1-67192270-furbici-planuji-ovladnout-lidstvo-rekla-hracka-z-90-let-pomoci-chatgpt [8.4.2023]

[189] Boj lidí s umělou inteligencí začal. Evropa se pokouší získat kontrolu nad ChatGPT https://archiv.hn.cz/c1-67192400-boj-lidi-s-umelou-inteligenci-zacal-evropa-se-pokousi-ziskat-kontrolu-nad-chatgpt [12.4.2023]

[190] Umělá inteligence ve školství: připraví na přijímačky, žáci s ní malují obrázky či píší básně https://archiv.hn.cz/c1-67191470-umela-inteligence-ve-skolstvi-pripravi-na-prijimacky-zaci-s-ni-maluji-obrazky-ci-pisi-basne [12.4.2023]

[191] Prohrává Evropa závod ve vývoji AI? Německý start-up věří, že se ChatGPT dokáže vyrovnat https://archiv.hn.cz/c1-67193560-prohrava-evropa-zavod-ve-vyvoji-ai-nemecky-start-up-veri-ze-se-chatgpt-dokaze-vyrovnat [13.4.2023]

[192] Týden s technologiemi: Umělá inteligence nalákala i Muska https://archiv.hn.cz/c1-67194010-tyden-s-technologiemi-umela-inteligence-nalakala-i-muska [14.4.2023]

[193] EU chystá tvrdou regulaci umělé inteligence. Europoslanci chtějí zakázat, aby software zkoumal emoce lidí https://archiv.hn.cz/c1-67195210-eu-chysta-tvrdou-regulaci-umele-inteligence-europoslanci-chteji-zakazat-aby-software-zkoumal-emoce-lidi [18.4.2023]

[194] Google pod tlakem umělé inteligence: Samsung chtěl nasadit v telefonech Bing https://archiv.hn.cz/c1-67194820-google-pod-tlakem-umele-inteligence-samsung-chtel-nasadit-v-telefonech-bing [18.4.2023]

[195] Musk chce vyvinout konkurenta pro ChatGPT. Jeho umělá inteligence bude hledat „maximální pravdu“ https://byznys.hn.cz/c1-67195370-musk-chce-vyvinout-konkurenta-pro-chatgpt-jeho-umela-inteligence-bude-hledat-maximalni-pravdu [18.4.2023]

[196] Microsoft testuje vlastní čipy pro umělou inteligenci. Mají nahradit drahé výpočetní karty od Nvidie https://archiv.hn.cz/c1-67195810-microsoft-testuje-vlastni-cipy-pro-umelou-inteligenci-maji-nahradit-drahe-karty-od-nvidie [20.4.2023]

[197] Muskův náročný týden: explozí rakety to neskončilo. Přišel v přepočtu téměř o 300 miliard korun https://byznys.hn.cz/c1-67197000-muskuv-narocny-tyden-explozi-rakety-to-neskoncilo-prisel-temer-o-300-miliard-korun [21.4.2023]

[198] Týden s technologiemi: Musk a Microsoft míří k soudu kvůli Twitteru a soupeření na poli AI https://archiv.hn.cz/c1-67196420-tyden-s-technologiemi-musk-a-microsoft-miri-k-soudu-kvuli-twitteru-a-soupereni-na-poli-ai [21.4.2023]

[199] Elon Musk vyvíjí svou TruthGPT. Kde končí pravda a začíná marketing? https://vikend.hn.cz/c1-67195920-elon-musk-vyviji-svou-truthgpt-kde-konci-pravda-a-zacina-marketing [21.4.2023]

[200] Jak může ChatGPT pomoci investorům? Odhadne pohyb akcií a výroky centrálních bankéřů přeloží do „lidštiny“ https://archiv.hn.cz/c1-67196470-jak-muze-chatgpt-pomoci-investorum-odhadne-pohyb-akcii-a-vyroky-centralnich-bankeru-prelozi-do-bdquo-lidstiny-ldquo [24.4.2023]

[201] Firmy se zapomínají učit z vlastních chyb, pak snáze podlehnou kyberútoku. Zločinci navíc získávají novou zbraň https://hn.cz/c1-67181710-firmy-se-zapominaji-ucit-z-vlastnich-chyb-pak-snaze-podlehnou-kyberutoku-zlocinci-navic-ziskavaji-novou-zbran [26.4.2023]