TECHNISCHE UNIVERSITÄT WIEN

Master Thesis

Socio-economic and Ethical Impact of Artificial Intelligence on Organizations and the Future of Work

A Thesis for the degree of

Diplom-Ingenieur (Dipl.-Ing.)

submitted at TU Wien

Faculty of Mechanical and Industrial Engineering

by

Ilya FAYNLEYB

Matr.No.: 01618896

Supervised by

Univ.Prof.in Mag.a rer.soc.oec. Dr.in rer.soc.oec.
Sabine Theresia Köszegi

(E330 - Institut für Managementwissenschaften)

Vienna, September 2024 _____

Contents

Abstract

The technological boom of Artificial Intelligence (AI) is having an extraordinary impact on organizations and the future of work at an unprecedented pace. The socio-economic and ethical implications of this technology are very significant and should not be overlooked. This study inspects the current state of AI, specifically generative AI, in relationship to ethical issues such as bias and discrimination, accountability, and worker autonomy, as well as socio-economic issues such as job displacement, upskilling and reskilling. The results show that ethical components of AI development are often neglected, which results in unreliable, unfair and biased AI that can already affect human lives. The socio-economic analysis shows not only the benefits of the adoption of generative AI, its improvement of workers' productivity and efficiency but also the effect it has on workers' cognitive skills, job displacement through automation, and the expected change of the most needed skills. Both socio-economic and ethical impacts are explored from the perspective of labour process theory, which is proposed as a framework for future research. As organizations and companies begin to increasingly rely on AI, there is a need to develop frameworks that balance technological advancement with ethical and socio-economic responsibilities. This master thesis provides a foundation for future research on a systematic approach to AI governance that addresses these impacts, ensuring that the benefits of AI can be kept, at the same time mitigating its potential risks to society and the future of work.

Chapter 1

Introduction

1.1 Problem Statement

As artificial intelligence (AI) continues to develop, its integration into our society has already begun. This integration creates unexpected and profound challenges, from both socio-economic and ethical points of view. Organizations and the future of work in general are on the verge of drastic changes, that will have an impact on millions of people (Acemoglu and Restrepo, 2018). While AI is expected to increase efficiency, productivity, and innovation within organizations, its adoption raises critical questions regarding income inequality, job displacement, and ethical decision-making. Making the situation even more alarming, the integration seems to accelerate, with disregard for possible implications (Acemoglu et al., 2022).

A particularly widespread and available to the public type of AI at this moment is generative AI. ChatGPT, Copilot, and Gemini are the most well-known and used examples of this technology. Despite being available to the public only since late November 2022, generative AI already had a great impact on our society not only on an individual but also on an organizational level. In some cases, integration of this technology has shown some improvements in efficiency and productivity and has promising potential for further implementation (Brynjolfsson et al., 2023).

This master thesis aims to define key concerns of the recent development and integration of generative AI and decision-making AI in general in organizations and investigates the socio-economic and ethical impact of AI on organizations and the future of work. Ultimately, this study aims to create a deeper understanding of the upcoming challenges and opportunities arising from the integration of AI technologies in organizations, as well as to create a foundation for more informed decision-making, policy development and AI governance regarding all AI technologies.

1.2 Objectives

In this master thesis, the current state of AI technologies will be reviewed and analyzed. This research aims to discuss the overall effects of AI on organizations and the future of work, however, its main objective is to answer the following questions:

1. How does the integration of AI technologies change cognitive skills and overall job requirements?
2. Which socio-economic and ethical implications does the widespread adoption of generative AI technologies within organizations have?
3. What recommendations can be given to the lawmakers and organizations that begin to implement AI?

Answering these questions can play a big role in understanding the current state of adoption of AI technologies, as well as identifying the flaws and possible improvements for future AI integration in different fields.

1.3 Methodology

For the analysis approach and the methodology, a systematic literature review (SLR) was taken as the basis. SLR aims to identify and evaluate all relevant literature on a topic to

derive conclusions about the question under consideration. This methodology was chosen because it can provide insights into the current state of research on a topic, at the same time identifying gaps and areas requiring further research about a given research question. The SLR approach also ensures a comprehensive, unbiased, and reproducible assessment of the existing literature related to the socio-economic and ethical impact of AI. The methodology follows the guidelines provided by Xiao and Watson (2019).

For this master thesis SLR process involved several key steps (Xiao and Watson, 2019):

1. *Literature identification:* Literature was found using key words using open sources like Google Scholar. Preliminary filtering was done based on the title and the limitation of the publication date.

2. *Screening for inclusion:* Abstracts of the publications were read to further filter only relevant sources that could be used for the study and to avoid irrelevancy.

3. *Quality and eligibility assessment:* Full texts were read to evaluate quality and eligibility. Only journal articles and books published by reputable publishers as high-quality research were included in this master thesis.

4. *Data extraction and analysis:* Finally, through the data extraction and analysis of the publications was the relevant information found and included in this study.

Following previously mentioned key steps, the research of the literature was conducted to comprehend the topic and to create a foundation for the analysis, discussion, and conclusions. With recommendations from the Institute of Management Science of TU Wien and individual research through Google Scholar, multiple research papers, journals, reports, and books, at a total number of 44, were thoroughly analyzed and used as the main source for this master thesis.

After reading and extracting the most important information from these sources, they were quoted and taken as the foundation for the theoretical background. Theoretical background was used to prepare the analysis of socio-economic impact and the following conclusions. Moreover, the evaluation of the socio economic impact of AI is based on labour process theory, that will be explained in later chapters.

Chapter 2

Theoretical Background

One of the most accessible to the public and popularized types of AI that already had strong influence on society, organizations, and individual workers in general is generative AI. Understanding this technology is essential for the analysis of the impact and implications that it had and will have in the future. This chapter's objective is to introduce the concept of generative AI and to provide important definitions related to the topic. Some examples of this technology will also be explored and analysed to create the background for future discussions.

2.1 Generative AI

Generative modeling artificial intelligence (GAI), or simply generative AI, is a machine learning framework, that generates man-made relics via the use of statistics or probabilities (Baidoo-Anu and Ansah, 2023), or a class of machine learning technology that can generate new types of content, such as text, images, video, and audio, by analyzing patterns in existing data (Brynjolfsson et al., 2023). The impressive aspect about this technology is its ability to identify patterns across enormous sets of data and, as previously mentioned, generate new content, which is something that has previously been considered uniquely human (Ellingrud, et al., 2023).

According to Brynjolfsson et al. (2023), recent progress in generative AI has been driven by

four main factors: computing power, earlier innovations in model architecture, the ability to ”pre-train” using large amounts of unlabeled data, and finally, refinements in training techniques. Different models of generative AI have different scales, computing power used for training, as well as model parameters and dataset sizes, which results in different models with different strengths, weaknesses, and applications

There are multiple sub-sections of this technology, however, due to the scope limitation and overall simplification and precision of the results, this research will focus on Large Language Models and Generative Pre-trained Transformers.

2.1.1 Large Language Models (LLMs)

Large Language Models (LLMs) are neural network models designed to process sequential data. Brynjolfsson et al. (2023) explain the main work principle of LLM as follows:

> “...LLM can be trained by giving it access to a large corpus of text (such as Wikipedia, digitized books, or portions of the Internet) and using that input text to learn to predict the next word in a sequence, given what has come before. This knowledge of the statistical co-occurrence of words allows it to generate new text that is grammatically correct and semantically meaningful.” (p. 4).

What makes LLMs different from the rest of the generative AIs is their model architecture based on two key innovations: positional encoding and self-attention. Positional encoding refers to the ability of the algorithm to keep track of the order in which a word occurs in a given output, which allows large amounts of input text to be broken in smaller segments that can be processed simultaneously. On the other side, self-attention refers to the ability of the algorithm to assign importance weights to each word in the context of the entire input text (Brynjolfsson et al., 2023).

Another important aspect of this model is its ability to be pre-trained on large amounts of unlabeled data, and because this kind of data is prevalent, this allows LLMs to learn about natural language on a much larger training corpus. Due to this aspect and since its training

is not specific to a particular set of tasks, LLMs can be used in multiple applications and also further ”fine-tuned” to match the priorities and needs of any specific setting or task (Brynjolfsson et al., 2023).

2.1.2 Generative Pre-trained Transformer (GPT)

Generative Pre-trained Transformer (GPT) can be considered a part of the LLMs family, however, there are a few differences. GPT models use publicly available digital content data to read and produce text in several languages (Natural Language Processing (NLP)) and can exhibit creativity in writing from a paragraph to a full research article convincingly on almost any topic (Baidoo-Anu and Ansah, 2023). As the name suggests, GPT models are based on transformer architecture, meaning that they are decoder-only models that use the same parameters to process the context and the source as a single input for generating the next output. GPT models are also characterized by their monolingual training data, predominantly English, and the necessity for a much larger number of parameters to achieve multilingual in-context capabilities (Hendy et al., 2023).

In summary, the main difference between the GPT model from the LLMs is that the latest model is designed to solve a wider range of tasks, however, is usually fine-tuned to a specific task for improved efficiency, while the GPT model excels at auto-regressive text generation. However, as the following research will show, the GPT model can be used in different ways and has diverse applications in a variety of fields.

The following sections will introduce some AI tools based on LLMs and GPT, which are already integrated by organizations and individuals on official or non-official levels.

2.1.3 ChatGPT

ChatGPT is a generative AI developed by the company OpenAI in 2022 and is one of the biggest and most popular generative AI tools with free public access to the basic version by the time of this research. For the sake of proving the capabilities of generative AI tools and showing how they can work, ChatGPT was asked the following question: “What is

ChatGPT?" Only after a few seconds, it provided the following answer:

> "ChatGPT is an AI language model developed by OpenAI, based on the GPT (Generative Pre-trained Transformer) architecture. It's designed to understand and generate human-like text based on the input it receives. ChatGPT can engage in conversations, answer questions, generate creative content, assist with tasks, and more, drawing from its training on a vast amount of text data from the internet. It's capable of understanding context, generating coherent responses, and adapting its language style to various topics and situations" (Generated by ChatGPT on May 5th, 2024).

To confirm, that the information of the tool is plausible, the official report of the latest version, GPT-4 OpenAI, (2023) was searched:

> "...GPT-4, a large multimodal model capable of processing image and text inputs and producing text outputs. Such models are an important area of study as they have the potential to be used in a wide range of applications, such as dialogue systems, text summarization, and machine translation" (p.1).

Comparing both statements, it does seem that the information provided by ChatGPT is valid, with some extra descriptions and examples.

ChatGPT is considered to be a so-called chatbot, a type of AI, that interacts with a user via chat to provide a better, more human-like and interactive experience.

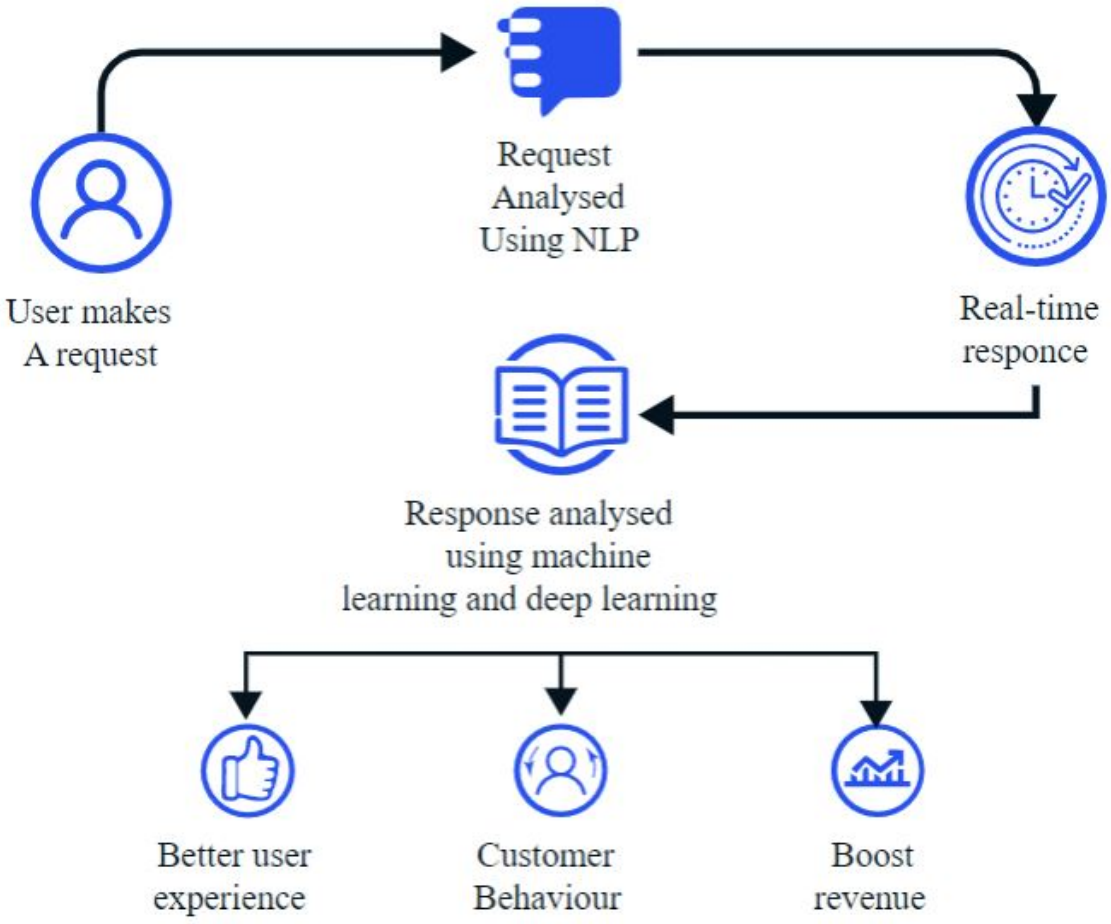


Figure 2.1: The architecture of a Chatbot (Gupta et al., 2023).

Figure 2.1 represents the basic architecture of a chatbot and its working principle: the user initiates requests, which are analyzed using NLP, and receives a real-time response from the chatbot. This response is then further analyzed to enhance the user experience in the subsequent conversation (Gupta et al., 2023).

By the time of the research for this master thesis, OpenAI has presented four versions of ChatGPT: GPT-3, GPT-3.5, GPT-4 and GPT-4o (openai.com, 2024). GPT-3.5 is the free version, that was used to provide previous information, whereas GPT-4 is the newest version, which is not free. The newest version provides much higher accuracy, as well as multiple other options and possibilities that the free version doesn't. For example, as the Figure 2.2 shows, performances in different exams of both versions of ChatGPT were

compared and GPT-4 dominates in almost all of them, exhibiting human-level performance and reaching the equivalent of the top 10 percent of test takers (OpenAI, 2023):

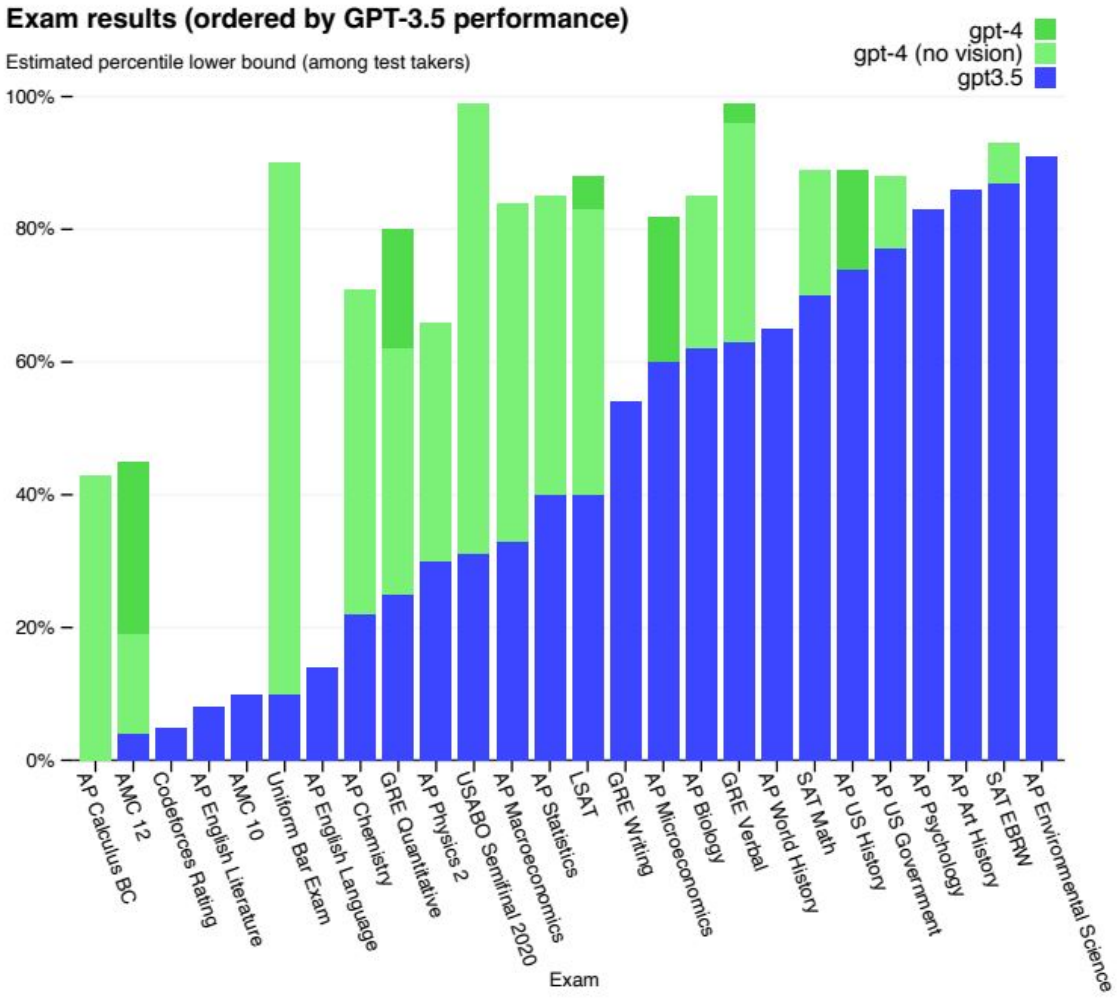


Figure 2.2: GPT performance on academic and professional exams (OpenAI, 2023).

Other areas that ChatGPT excels at are image recognition and processing, overall text generation and even problem understanding and solving, allowing it to be used to write code, design products, create marketing content and strategies, streamline operations, analyze legal documents, provide customer service via chatbots, and even accelerate scientific discovery (OpenAI, 2023; Ellingrud et al., 2023). All these advantages and assets create a very powerful tool that is highly attractive to different stakeholders. Seeing the possibilities and the potential of ChatGPT it is no wonder, that individuals, as well as organizations,

are interested in its implementation for work acceleration and also improved productivity and efficiency.

2.1.4 Microsoft Copilot

Microsoft Copilot is another example of a new AI tool that is being quickly integrated into the market. Based on LLMs and specifically the GPT model from OpenAI, Microsoft Copilot is an AI tool that serves as an advanced AI assistant designed to augment human productivity across various tasks (Adetayo et al., 2024).

What makes this tool different from ChatGPT is its integration in various Microsoft products such as the operating system Windows 11, as well as Bing Search, Microsoft Edge and Microsoft 365 apps like Word, Excel, PowerPoint, Outlook and Teams (Microsoft, 2024). When it comes to its capabilities, Microsoft Copilot has lots of advantages compared to the other AI tools. Quoting Adetayo et al. (2024):

> "Microsoft Copilot is a versatile communication tool that excels in managing diverse conversations. It supports multiple languages, adapts to various styles and offers advanced contextual awareness. Notably, it suggests complete responses, streamlining collaboration and fostering efficient teamwork, especially in projects with changing specifications. ... Copilot offers seamless integration with desktop applications." (p. 1).

This integration of the tool directly into the Microsoft services, combined with its versatile functionality, makes this AI assistant especially attractive to organizations and office workers who already use Microsoft's products and services.

2.2 Ethical Guidelines of AI

The impact of AI is not limited to the socio-economic implications. As already mentioned in previous sections, currently AI is an imperfect technology that affects thousands of lives

and is only expected to expand its influence to more professional fields and occupations. This impact has to be considered not only from points of view of productivity, efficiency and economic viability but also from the perspective of ethics for a better understanding of existing challenges, as well as the challenges that are about to come. Understanding ethical challenges is essential for the further development of any technology and it is especially important in the case of AI tools that are being integrated into society at extreme rates. Addressing and working on these challenges may have a positive impact on the existing state of technology and smoothen out the negative impact that is already present in the current state of implementation of AI.

This chapter will explore current concerns of AI technologies such as bias and discrimination, privacy, transparency and accountability and finally, decision-making processes and autonomy.

2.2.1 Bias and Discrimination

Various concerns and problems are often mentioned and addressed when talking about the ethical implications of AI, however, currently, bias and discrimination are some of the most prevailing.

The UNI Global Union report (2017) describes bias as the action of using features such as gender, race, sexual orientation, and others as discriminatory elements in a decision with a negative impact somehow harmful to the human being. Varona and Suárez (2022) further explore the connection of definitions of bias and discrimination and shortly summarize as follows:

> "Discrimination and bias are two entangled variables with a strong interdependency that results in one of them being the cause and the effect of the other. ... bias refers to the action of deciding upon an individual or group with a given potentially harmful impact because of their features, while discrimination is expressed by the outcome of the decision itself" (p. 11).

The use of biased AI has numerous ethical implications that must be carefully considered.

Biased AI has a strong potential for discrimination against individuals or groups based on factors such as race, gender, age, or disability, amplifying existing inequalities and reinforcing discrimination against marginalized groups (Ferrara, 2023).

Specifically in AI technologies, biases are presented in different forms and have different causes. Ferrer et al. (2022) identify three causes that have been distinguished:

1. *Bias in modeling:* Can be manifested in form of so-called algorithmic processing bias, where biases are deliberately introduced through smoothing or regularization parameters to compensate for bias in the data.

2. *Bias in training:* A more common "infusion" of biases happens in this stage, where the algorithm learns to make decisions or predictions based on data sets that already reflect existing prejudices. This makes the algorithm more likely to learn to make the same biased decisions.

3. *Bias in usage:* When the algorithm is used in a situation for each it was not designed, it often also results in a manifestation of biases. An algorithm utilized to predict a particular outcome in a given population can lead to inaccurate results when applied to a different population. This phenomenon is called a transfer context bias.

The process of automated decision-making encapsulates the emergence of discrimination as the effect of bias in training:

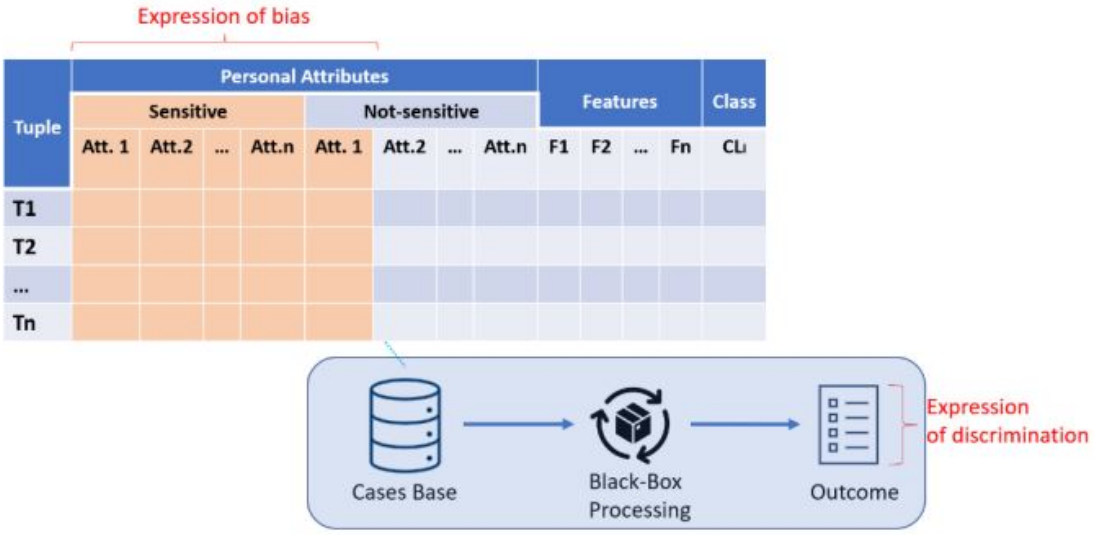


Figure 2.3: Bias and discrimination in an automated decision-making process (Varona and Suárez, 2022).

The effects of the bias that are present in AI, similar to human prejudice, lead to discriminatory predictions and recommendations. Figure 2.3 describes how biases can be expressed in the inclusion of a subset of attributes (sensitive and non-sensitive) oriented to the subject identification from the set of attributes describing a particular individual. Discriminatory outcome is the result of the consideration of these attributes in the decision (Varona and Suárez, 2022). This model is a great representation of the idea of Ferrer et al. (2022) of how biases in training are established.

Varona and Suárez (2022) also provide the following statement about the appearance of the bias in the training data:

> "...while learning, systems are designed to spot patterns, and if the training data is unrepresentative, then the resulting identified patterns will reflect those same patterns of prejudice and, consequently, they will produce unrepresentative or discriminatory decisions/predictions as well" (p. 5).

Therefore, it is fair to say, that if repeated discriminatory or biased conclusions are present, there is an existence of an unfair dataset and that any instance of an algorithm using that

dataset for training will produce equally unfair future decisions and predictions (Varona and Suárez, 2022; Ferrer et al., 2022).

The existence of such biases can easily be found in most generative AI models. For example, when models are prompted to create images of CEOs, they tend to reinforce stereotypes by depicting CEOs predominantly as men. At the same time criminals and terrorists are portrayed as people of color, leading to new forms of discrimination, such as those based on skin color, ethnicity, or even physical appearance (Ferrara, 2023).

The fact that AI technologies are not exempt from biases, such as race-based or sex-based, is very often overlooked by organizations and companies when applying AI-based skill development strategies. This negligence can often result in counter-productivity and exacerbation of existing inequalities within the organization and society at large. Thus, organizations must carefully manage the implementation of AI and consider its impact on learning and development and ensure that any strategies they implement take into account the diverse needs and perspectives of their workforce (Morandini et al., 2023). However, it is also the responsibility of developers, companies, and governments to ensure that AI systems are designed and used fairly and transparently (Ferrara, 2023).

2.2.2 Privacy Concerns and Data Collection

Another major concern regarding AI technologies is the potential misuse of private data. The social implications of this issue should not be underestimated, particularly in light of the rapid, and as previously mentioned, mostly uncontrolled, advancements of AI development. Carmody et al. (2021) point out that these advancements, along with the vast amounts of personal information available online and in the public domain, combined with the Internet of Things (IoT) and social media, have greatly increased the overall number of privacy violations. Modern surveillance practices have become commonplace in society. In recent years there have been numerous instances where companies have collected personal and private data for secondary profits or material gain. This data has often been shared with third parties, either intentionally through sales or unintentionally due to hacking or accidental privacy breaches. Nevertheless, the increasing collection and deliberate dissem-

ination of customer data to third parties have become a major privacy concern (Carmody et al., 2021).

This issue especially impacts human resource management and related technologies, particularly machine learning. Many tools developed to automate and expedite hiring processes require substantial data to train and predict machine learning models, however, it is often neglected that this data often includes various amounts of sensitive or private information (Faynleyb, 2024).

When it comes to generative AI, there are also various concerns regarding privacy and security. For example, generative AI tools such as ChatGPT have shown a double-edge nature in the realm of cybersecurity, since there were documented several instances of the use of AI in both defensive and offensive sides of cybersecurity (Gupta et al., 2023). On one side, generative AI tools can aid cyber defenders by leveraging large LLMs trained on extensive cyber threat intelligence data, which helps in enhancing threat intelligence capabilities by extracting insights and identifying emerging threats.

Quoting Gupta et al. (2023):

> "The GenAI tools can also be used to analyze the large volume of log files, system output, or network traffic data in case of cyber incidence. This allows defenders to speed up and automate the incident response process. GenAI driven models are also helpful in creating a security-aware human behavior by training the people for growing sophisticated attacks. GenAI tools can also aid in secured coding practices, both by generating the secure codes and producing test cases to confirm the security of written code. Additionally, LLM models are also helpful to develop better ethical guidelines to strengthen the cyber defense within a system" (p. 3).

On the other side, due to the public nature of the GenAI tools and their free access, the same benefits mentioned for cyber defenders can also be used for cyber attacks by the cyber offenders. Cyber offenders can exploit generative AI tools to perform more effective cyber attacks by extracting information or bypassing ethical policies of tools like ChatGPT

or use them to create convincing social engineering and phishing attacks, generate attack payloads, and produce various types of malicious code. Making things worse, the ease of access and usage of tools such as ChatGPT allows even those with little technical knowledge of programming and overall IT security to conduct sophisticated cyber attacks (Gupta et al., 2023).

Overall, regarding privacy and data collection, several points must be addressed. First of all, LLMs' use of personal information for training and responses can conflict with the European Union's GDPR compliance laws and to fix this, the developer needs to discuss and ensure that the LLM adheres to those laws, as LLMs could potentially be banned from those countries if not (Gupta et al., 2023). Another point is that the sensitive information provided to and kept by the AI tools, is also a concern for lots of people (Allhutter et al., 2018; Gupta et al., 2023), however, if LLMs simply do not save a user's chat history, company policies, or have the option to delete messages from the LLM's history might be a possible the solution to the problem.

LLMs also suffer from the data collection problem: new and existing models should be continuously trained and updated to prevent outdated information from being given to the user; ChatGPT's information cutoff being September 2021 is a good example of this limitation. Updating and training models is critical for the proper and ethical work of AI tools since there is likely to be more old than new information on a given topic, which could cause the model to place more trust in the old information, making it biased and discriminating, as also discussed in the previous section (Gupta et al., 2023).

Another problem regarding data collection and its quality is discussed in the paper of del Rio-Chanona et al. (2023), where the problem of training LLMs with AI-generated data and the overall decline of human-generated content is addressed. The authors point out, that ChatGPT was adopted at an unprecedented rate, causing a significant decrease in content creation on Stack Overflow (a knowledge repository for programmers) after ChatGPT's release, indicating that users are substituting Stack Overflow with ChatGPT for their programming questions. The problem with this decline is the existence of a risk to digital public goods, which are essential for the training of future AI models. Since LLM-generated content is not a reliable substitute for human-generated data, training on

LLM-generated content can lead to progressively poorer results. This process is further enhanced by the shift of the data from open to closed access, where data is controlled by private companies like OpenAI. This trend could prevent future development of AI models and reduce the availability of public knowledge (del Rio-Chanona et al., 2023).

Finally, just like the study of Gupta et al. (2023), the study of del Rio-Chanona et al. (2023) highlights the potential for increased inequality, since leading AI models gain exclusive access to valuable user interactions, giving them a competitive edge in comparison to the rest of the competition on the market or even countries that don't have access to the technology. In the end, the consolidation of information-seeking behavior around a few LLMs could narrow the diversity of information and limit exposure to new ideas and concepts.

2.2.3 Transparency and Accountability

Another concern that is very often addressed when analyzing AI technologies, is the problem of transparency and accountability. The definition of transparency can be analyzed from the perspective of a physical property of a material, that allows the light to go through the material, however, in the case of AI, it goes one step further and gets a more metaphorical meaning, mostly related to "seeing". The light is associated with brightness and clarity and therefore with transparency. Several neighboring concepts of particular relevance for transparency that relate to AI exist, but they can be simply summarized as "explainability" and "openness" (Larsson and Heintz, 2020).

According to the study by Kim et al. (2020), current state-of-the-art AI technologies are based on the application of deep learning, especially Convolutional Neural Networks (CNNs) and Recursive Neural Networks (RNNs), which have revolutionized the NLP field. The models, however, are very often criticized for their lack of interpretability and transparency, leading to hesitance in their use for decision-making in critical tasks. This complexity is also described as the "black box" problem, which obstructs the transparency and accountability of most AI systems (Kim et al., 2020).

Quoting Faynleyb (2024):

> "The concept of transparency is especially important, since in recent years more and more decision-making processes are being transferred from humans to AI. Therefore, making AI technologies more interpretable by providing appropriate explanations is a fundamental step in building trust with the users, since without this trust, the users will not be willing to accept the solution. To address these concerns, recent legislative and regulatory trends are increasingly demanding fairness and transparency in algorithms, forcing explanations to be more mandatory than preferable. Based on an "explainable" AI, it is possible to build AI systems that can be understood, trusted, and effectively managed by humans" (p. 10).

Therefore there is a strong necessity for an explainable AI, that is transparent and fair, for future development, and Kim et al. (2020) propose such a framework.

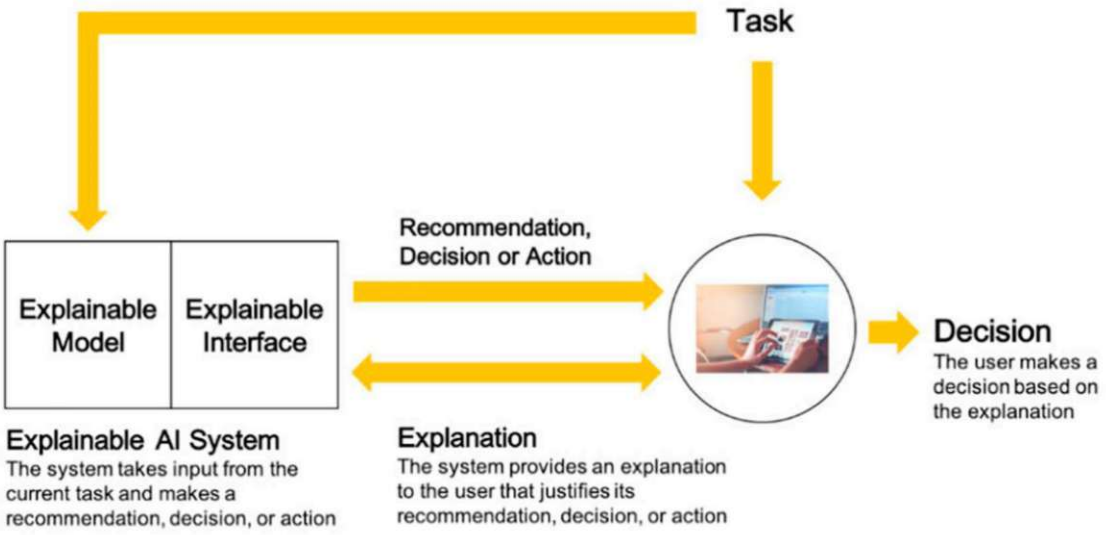


Figure 2.4: Explainable AI Framework (Kim et al., 2020).

The framework seen in Figure 2.4 portrays the workflow of the tasks that enter the explainable AI system. Just like in a normal AI framework, an explainable AI system provides the output to the task in the form of a recommendation, decision or action, however, most importantly, an explanation of the decision made can be requested. The system should be

able to justify its own decision, recommendation, or action, and based on this explanation the user can make the final decision, that is more transparent.

As mentioned by Faynleyb (2024), this framework is one way of providing a transparent model, that is easily understood and most importantly can be trusted in the decisions that it provides, solving the "black box" problem. Akinrinola et al. (2024) further support this idea by stating that techniques such as decision trees, rule-based systems, and model-agnostic approaches also contribute to the interpretability of AI systems. However, even with the issue of transparency out of the way, the issue of accountability still remains. There are three key strategies, that may help to solve this issue: regulatory measures, ethical AI governance, and human-in-the-loop approaches (Akinrinola et al., 2024).

Regulatory measures refer to the legal regulations provided by the governments of countries or organizations like the United Nations. With the development of AI, the regulatory landscape has also evolved, with many countries and regions considering or already implementing measures to govern AI development and deployment. Nevertheless, it is important to notice, that the applicable AI regulations are still in nascent stages. Making things worse, the staggering speed of AI development makes it difficult for the existing regulations to keep up. Akinrinola et al. (2024) also have stated the following about this strategy:

> "Understanding the limitations of current regulations highlights the need for a proactive approach to address ethical considerations beyond legal requirements. Proposals for strengthening AI regulations revolve around the development of comprehensive frameworks that address ethical considerations, accountability, and transparency. The aim is to create a regulatory environment that guides AI developers, ensures user protection, and establishes mechanisms for accountability in case of misuse or adverse consequences" (p. 53).

Therefore, this strategy requires a strong collaboration between policymakers, engineers and ethicists to provide the most up-to-date regulations and laws to ensure accountability in AI development and usage. One of the examples of such initiatives is the European Commission's Artificial Intelligence Act, where the requirements for high-risk AI systems,

conformity assessments, and penalties for non-compliance are outlined (Akinrinola et al., 2024).

The second approach, ethical AI governance, refers to the establishment of committees or boards, which are dedicated to overseeing the ethical aspects of AI development, but on a smaller scale, within organizations. These committees provide ethical guidelines to AI developers and organizations and should be inclusive, representing a wide range of stakeholders to foster a holistic approach to ethical decision-making. The guidelines should promote transparency, accountability, and user privacy to avoid social issues such as bias and discrimination (Akinrinola et al., 2024).

Finally, the human-in-the-loop (HITL) approach directly incorporates human oversight into AI systems, especially in critical decision-making processes. This approach is essential in scenarios where AI decisions have significant societal impact, such as healthcare, criminal justice, and autonomous systems. Akinrinola et al. (2024) highlight the following purpose of introducing HITL:

> "This ensures that human judgment is an integral part of the decision loop, allowing for the correction of potential biases, ethical considerations, and complex contextual understanding that AI systems may lack. HITL approaches strike a balance between automation and human intervention, mitigating risks associated with fully autonomous systems" (p. 54).

The involvement of humans in automated decision-making processes is crucial, since the existing risks of unsupervised autonomous systems are still very high, as it has already proved to be, especially in the sector of human resource management and bureaucracy, where human lives play a direct role (Faynleyb, 2024). The topic of automated decision-making will be analyzed in more detail in the following section.

2.2.4 Ethical Automated Decision-making

All of the above-mentioned concerns are either directly or indirectly involved in the process of decision-making. Before the appearance of AI and even computers, which could have

automated or overtaken the task of decision-making, this responsibility was solely human and the ethical issues such as biases, discrimination and accountability were also related to humans only. However, with the appearance and integration of information communication technology (ICT), decision-making processes began to be increasingly more automatized, blurring the line of accountability and ethical responsibilities between humans and machines (Faynleyb, 2024).

To begin exploring the topic of automated decision-making, it is important to understand the meaning of the concept itself. Araujo et al. (2020) describe the automated decision-making process as follows:

> "...it may be seen as the process through which the ever-growing amount — and variety — of personal data are subsequently processed by algorithms, which are then used to make (data-driven) decisions" (p. 611).

According to this definition, an automated decision-maker can be understood as an algorithm, a recommender system, or simply AI depending on its framing and presentation to the system's user or the decision's subject. Human involvement is limited and varies with the type of algorithm. The algorithms can range from decision-support systems that offer recommendations to human decision-makers to fully automated decision-making processes that operate independently on behalf of institutions or organizations (Araujo et al., 2020).

The idea of such algorithms has appeared already half a century ago, in the early 1970s. Back in the day, the algorithm was seen as a so-called "decision support" system, which was designed to help managers report, analyze, and interpret data, that would consequently be used by the manager to make the final decision (Harris and Davenport, 2005). Nowadays, however, through the technological advancements and accelerating development of AI, the algorithms are often exceeding mere decision support and often replace humans entirely, completely avoiding human oversight. This negligence of human involvement in the automated decision-making processes is one of the main causes of the ethical issues that are often mentioned when talking about AI development (Faynleyb, 2024).

With the appearance of generative AI, the capabilities and even the role of AI had to

be reevaluated. Whereas in the past AI was mostly imagined to be analytic, suitable for decision-making tasks, nowadays it gains the capability to perform generative tasks, even suitable for content creation, and creativity itself, something unthinkable back in the day (Feuerriegel et al., 2024). The responsibility and accountability of generative AI, however, is often a space for debates, since it does not make any decisions itself, but it can influence the final decision of the user.

The interaction of the user with generative AI nevertheless creates a new form of relationship, which Feuerriegel et al. (2024) address in the following statement:

> "With generative AI, a human uses prompts to engage with an AI system to create content, and the AI then interprets the human's intentions and provides feedback to presuppose further prompts. At first glance, this seems to follow a delegation pattern as well. Yet, the subsequent process does not, as the output of the AI can be suggestive to the other and will inform their further involvement directly or subconsciously. Thus, the process of creation rather follows a co-creation pattern, that is, the practice of collaborating in different roles to align and offer diverse insights to guide a design process" (p. 116).

This interaction of human-algorithm creates a so-called hybrid intelligence, which addresses each one's limitations and strives to combine human intuition, creativity, and empathy with the computational power, accuracy, and scalability of AI systems to achieve enhanced decision-making and problem-solving capabilities (Feuerriegel et al., 2024).

Even at the existing level of development of AI and automated decision-making, there already are some risks that are often neglected or simply not considered. The study by Araujo et al. (2020) provides some of the main findings regarding the existing attitudes toward automated decision-making.

First of all, the overall perception of objectivity and fairness is often associated with automated decision-making: there is a general belief that automated systems are more objective and rational compared to human decision-makers. This belief, the so-called "machine heuristic", stems from the assumption that statistical and algorithmic methods outperform

human judgment, which leads to a preference for algorithmic decisions over human ones.

Second, despite concerns about biases, individuals tend to have a cautiously optimistic view of automated decision-making. This optimism is even more pronounced among those who are typically critical or concerned about biases and discrimination in human decision-making. Other factors such as knowledge, online privacy concerns, demographics, and belief in equality also influence general attitudes towards automated decision-making.

Another interesting factor that changes the attitude of individuals towards automated decision-making is the context of its use: when examining specific contexts such as media, public health, and justice, perceptions of automated decision-makers versus human decision-makers show remarkable variations: in higher impact decisions such as justice and health, algorithms are often viewed more positively than human decision-makers in terms of fairness and usefulness, with lower perceived risks, whereas in lower impact decisions there are fewer differences, although human experts in justice are seen as marginally better than the algorithms.

All of these findings provided by Araujo et al. (2020), despite being optimistic towards the algorithm provide a very important insight to the general assumption of individuals about the algorithms: its assumed "neutrality". Quoting Araujo et al. (2020) regarding the nature of algorithms and automated decision-making:

> "...they are created for purposes that are often far from neutral: to create value and capital; to nudge behavior and structure preferences in a certain way; and to identify, sort and classify people" (p. 613).

Considering this statement, the overall understanding and attitude of people towards automated decision-making, especially ones unaware of its nature and principle of work, becomes at least incomplete and misleading.

It is clear, that the interaction between humans and AI is very new and needs further research. Furthermore, to explain and guide the behavior of humans who work with AI, the establishment of human-AI interaction models will be necessary to guarantee the effective, efficient and ethical use of AI (Feuerriegel et al., 2024).

2.2.5 Autonomy

Finally, all of the above mentioned ethical issues result in a direct impact on the person's autonomy. According to Rubel et al. (2020), autonomy is composed of self-rule, self-governance, or self-determination:

> "The ability to self-govern includes the ability to develop one's own conception of value and sense of what matters, to [develop] the values that will guide one's actions and decisions, and to make important decisions about one's life according to those values where one sees fit" (p. 550).

As the authors mention, the concept of self-governance is based on the values developed by a person with life experience. Therefore it is often frowned upon, when an autonomous algorithm participates or even completely replaces humans in decision-making processes, undermining human autonomy. However, human autonomy is a very complex concept, that can be analyzed in different ways.

Laitinen and Sahlgren (2021) present a multi-dimensional model that describes multiple aspects of human autonomy. The model consists of several components: capacities and requirements, respect, exercise and resources. The first aspect refers to the requirement of certain capacities that enable individuals to make their own decisions and act according to them. These capacities include cognitive abilities to understand and process information, moral competence to understand ethical principles and apply them in decision-making and finally, social skills as the ability to communicate and interact effectively with others. These capacities form the foundation of self-determination, which is essential for the reflection on the values and goals. Another important part of the normative requirements in regards to autonomy, is the appearance of duties and corresponding rights (Laitinen and Sahlgren, 2021).

Other aspects such as respect and self-respect are also essential facets of autonomy. Quoting Laitinen and Sahlegren (2021):

> "The importance of interpersonal respect or recognition is closely tied to the fact

> that humans are born as merely potentially autonomous persons and need recognition and respect to develop their capacity for self-determination. It is clearly wrong and discriminatory to systematically block some people (due to their gender, "race", or caste) from developing their capacity for a self-determined life. When others respond by recognizing the person as autonomous and respecting them, a relational aspect of autonomy is formed" (p. 4).

This statement highlights how interpersonal relationships and mutual respect enhance the autonomy of an individual and how autonomy itself is a social construct in itself. Self-respect, on the other hand, refers to the relationship of the individual with himself and, despite the individual being the only participant of the relationship, can still be considered a social construct, since self-relations can be enhanced by recognition from others (Laitinen and Sahlgren, 2021).

The final components of the model are the exercise and the resources, which refer to the actual implementation or exercise of autonomy through action and decision-making and the conditions or resources to do so. The main idea of an autonomous exercise is independence from a heteronomous lifestyle, a state where an individual is governed by something other than himself and includes blind obedience to tradition and authority without forming an independent view of one's own. The real implementation of autonomy, however, is free from any external governance and is only responsible for the individual's own values and the real self. To achieve the exercise, the conditions must be met: from material and economic resources to cultural and informational prerequisites, which are important components of autonomy (Laitinen and Sahlgren, 2021).

Now, after presenting the multi-dimensional model, the question arises: how do AI and automation influence autonomy? Despite being created for the purpose of assisting and supporting people, AI systems can also be a hindrance to them. It appears that AI presents multiple kinds of obstacles to users' autonomy that restrict the development or exercise of individuals' capacities for self-determination. While AI systems themselves are not moral agents and cannot literally disrespect human autonomy, they are governed by "ought-to-be-norms," which dictate how AI should function to respect human autonomy. Laitinen and

Sahlegren (2021) continue expanding on this topic, presenting several forms of interpersonal disrespect that AI systems present.

The most apparent and influential form of disrespect is direct interference, when AI systems physically prevent an individual from performing an action, for example, by locking a user out of certain functionalities without consent, thus directly interfering with the user's autonomy. Therefore, it is essential to design the algorithms in a way that prevents such extreme interference (Laitinen and Sahlgren, 2021).

Other, more subtle, and much more relevant forms of disrespect presented specifically in generative AI are coercion, manipulation, and deception. In the context of AI, coercion involves forcing the user to make specific choices by removing meaningful options or offering options the user can not refuse, while manipulation and deception influence individuals' decisions by creating false beliefs or exploiting vulnerabilities (Laitinen and Sahlgren, 2021). The latter is a specifically acute issue of modern generative AI such as ChatGPT, which, as previously discussed, still suffers from biases, discrimination, and overall "hallucinations" that many users take at face value without questioning the legitimacy of the provided information. Such forms of involvement and influence strongly impact human autonomy and should be addressed and corrected in the development stages to create meaningful choice alternatives in a transparent manner.

Some other forms of discrimination that Laitinen and Sahlegren (2021) also mention in their paper are nudging and paternalism. Nudging refers to the subtle guidance of individuals toward certain decisions, while paternalism involves making decisions on behalf of individuals. The question arises: when, if ever, should the system make such involvements that interfere with an individual's autonomy? Beauchamp (2008) describes the following conditions to justify such interference:

1. *The Harm Condition:* A person is at risk of a substantial and preventable harm or loss of a benefit.

2. *The Likelihood Condition:* The paternalistic action has a strong likelihood of preventing the harm or obtaining the benefit.

3. *The Weight Condition:* The projected benefits of the paternalistic action outweigh its risks.

4. *The Minimal Interference Condition:* The least autonomy-restrictive alternative that will secure the benefits or reduce the risks is implemented.

These conditions are a good guideline for the overall design of automated decision-making for a more transparent and ethical interaction between the user and the algorithm.

Overall, Laitinen and Sahlegren (2021) conclude that the design of AI systems should enhance and support human autonomy rather than undermine it. One way to do so is by adhering to the "ought-to-be-norms" that protect and promote human self-determination and avoid interference, manipulation, and deception.

Aside from the autonomy of a user, the topic of autonomy can also be addressed to the AI itself. Considering the provided definition of autonomy and self-governance, it is understandable why lots of scientists and science fiction writers scrutinized and criticized the idea of an independent, autonomous AI. Some of the most well-known science fiction writers, Isaac Asimov and Philip K. Dick have written several novels on this topic. In their works "I, Robot" and "Do Androids Dream of Electric Sheep" accordingly, they focus on the topic of robots, AI, and the impact and influence of such technologies on humankind, robots themselves, and the interaction between them. They show the complexity of AI and how very often humanity doesn't even understand its own creations, which is very relevant to the current state of development of AI technologies. For example, one of the most revolutionary and new ideas from Isaac Asimov's books is the so-called "Three Laws of Robotics," first introduced in his 1942 short story "Runaround," which establishes the behavior of all robots:

1. *First Law:* A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. *Second Law:* A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. *Third Law:* A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Despite being very well and clearly defined, based on pure logic to determine a precise output, the actions of the robots in the stories are very often contradicting, unexpected and surprising. Robots often show signs of consciousness, self-awareness, and autonomy unforeseen by anyone.

Such unanticipated events, even though in a different, more primitive way, are observed nowadays with the development of all the new AI technologies, when people are asking ethical questions about the legitimacy of AI in its decisions and its role in human society. Reading and analyzing hypothetical and fictional stories, such as the ones of Isaac Asimov, can be of great contribution to the creation and further development of technologies in a more ethical way.

2.3 Labour Process Theory and AI

There is a need to view and analyze all of the above-mentioned issues and manifestations of AI in a single framework, to create a comprehensive model of relationships between AI and affected systems or stakeholders. One such framework can be the labour process theory (LPT), and this section's objective is to explain, why it may be a good way to observe and analyze the impact of AI on socio-economic, as well as ethical issues.

To explain the LPT, the definition of labour should be provided first. According to Karl Marx (1976), labour can be defined as the following:

> "Labour is, in the first place, a process in which both man and Nature participate, and in which man of his own accord starts, regulates, and controls the material re-actions between himself and Nature. He opposes himself to Nature as one of her own forces, setting in motion arms and legs, head and hands, the natural forces of his body, in order to appropriate Nature's productions in a form adapted to his own wants. By thus acting on the external world and

changing it, he at the same time changes his own nature" (p. 127).

From this definition, it can be implied, that labour is the process of interaction between an individual and nature, in which a material transformation occurs.

With the provided definition of labour, the concept of LPT can be summarized as a critical framework for understanding the dynamics of work and labour within capitalist systems, which emphasizes the control and organization of work processes (Knights and Willmott, 2016). On a more detailed level, LPT also analyzes the 'conversion movement' that transforms labour power into a commodity (Gandini, 2019). The reason, why LPT is a suitable framework for the analysis of AI impact, is the possibility to gain insights into how technological advancements influence labour control, worker skill development, and worker autonomy.

One example, which is essential in the context of LPT, is job displacement and job creation due to AI integration. LPT also explores how technological changes can lead to job displacement and shifts in employment patterns, therefore directly coinciding with the topic of AI impact. This way, LPT can help to determine and explain the net effect on employment due to jobs created or replaced by automation.

Another important concept of LPT according to Gandini (2019) is the deskilling or 'degradation' of labour, which focuses on complex tasks being broken down into simpler, repetitive ones that require less skill and are easier to control. As multiple authors have highlighted (Li, 2022; Morandini et al., 2023; Ellingrud et al., 2023), AI algorithms can automate routine and repetitive tasks, potentially leading to deskilling or even the full disappearance of jobs in certain sectors. At the same time, as will be shown later, they also create opportunities for reskilling and upskilling of workers as the workforce needs to adapt to new technologies and new roles that require higher levels of digital literacy and technical expertise. These changes and trends fit well in LPT and can also be further explored and analyzed with it.

As also explored in the autonomy section, AI tools have a very strong impact on worker's autonomy and it is also a major concept of LPT since LPT considers how changes in the labour process can impact worker autonomy and empowerment.

To conclude this section, it should be stated, that LPT has great potential for the analysis of the impact of AI and it can be suggested as a framework for further research on the topic. Even though this theory may have been criticized in the past by various researchers (Gandini, 2019), the development of digital technologies has shown LPT to be very suitable for observation and contribution to the ethical AI debate, and established as a critical Marxist voice on this topic. This theory will be used for the analysis of the socio-economic impact that will be explored in the following chapter.

Chapter 3

Socio-economic Impact

Like any other revolutionary technology, AI and automation in general are expected to have a strong impact on modern society. From industrial robots to automated document processing systems, these technologies continue to be the biggest factor in changing the demand for various occupations, and generative AI is a great representation of such technologies. In recent years generative AI has accelerated automation and extended it to a new set of occupations, amplifying the impact (Ellingrud et al., 2023).

This chapter will cover some of the most significant socio-economic topics such as job displacement, digitalization, and economic growth, in relation to automated technologies, specifically generative AI.

3.1 Job Displacement vs. Job Creation

One of the major concerns regarding the integration of automation is work redundancy. Some studies argue, that as digital technologies, robotics, and AI are being integrated into various aspects of society, workers will find it increasingly difficult to compete against machines, and their compensation will experience a relative or even absolute decline (Acemoglu and Restrepo, 2018). For some, automation in its current form may even be the first sign of a jobless future, while others consider it as enriching human productivity and work experience (Acemoglu et al, 2022). In either case, the integration of automated technologies is

going to have an impact on modern society. It is the question of the speed, the magnitude of the impact and the depth of the integration.

AI and ICT in general seem to have a very large influence on the development of job displacement and creation. Digital skills appear to improve the individual's labor market opportunities, and hence the occupation's tendency to decline or increment (Chen et al. 2022). In recent years AI development has changed the essence of work, making ICT skills a fundamental requirement of the modern labor force and at the same time transforming occupations, making them attract employees with digital skills to adapt to the increasingly digital environment (Chen et al. 2022). With current technological advancements, some people are not able to get good jobs due to a lack of the right skill set, while others are afraid of low-skilled jobs being threatened by automation (Li, 2022).

A recent study by consulting company McKinsey (Ellingrud et al., 2023) suggests that automation, and especially the recent development of generative AI, will sharply accelerate and affect a wider set of work activities in the US, involving expertise, interaction with people, and creativity. As can be seen in Figure 3.1, it was estimated, that without generative AI, automation could take over tasks accounting for 21,5 percent of the hours worked in all the sectors of the US economy by 2030, however with it, that share has increased to 29,5 percent (Ellingrud et al., 2023):

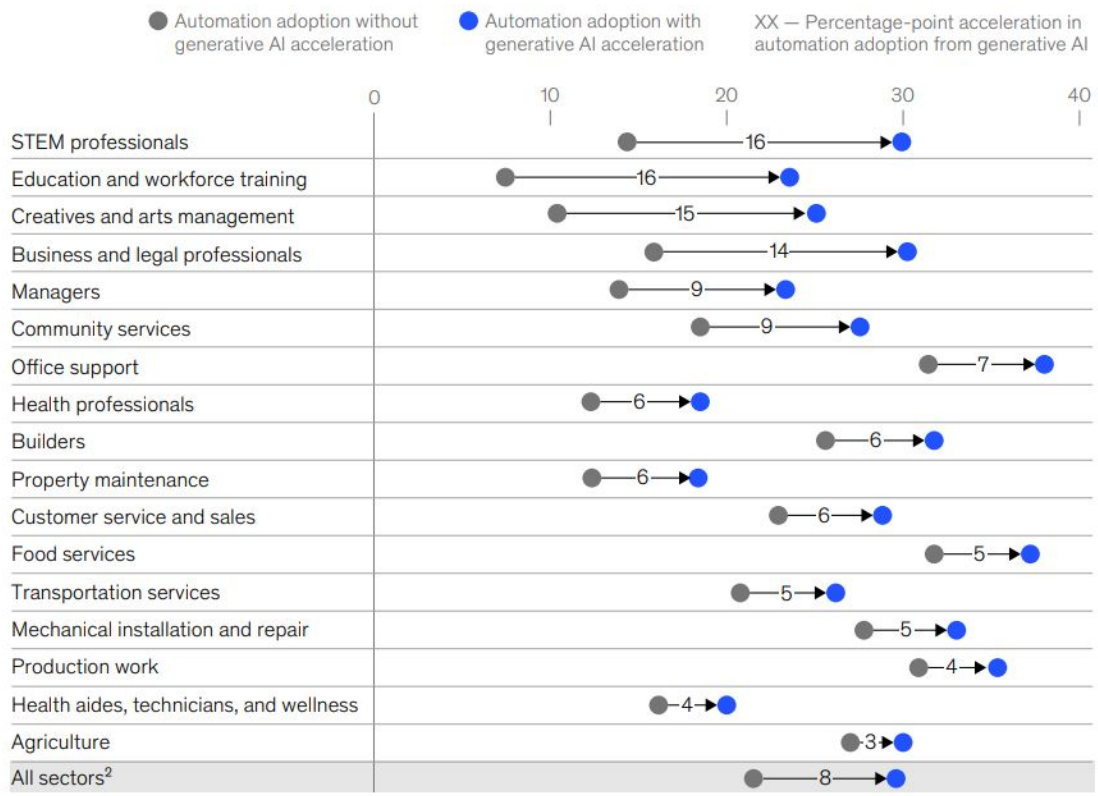


Figure 3.1: Midpoint automation adoption by 2030 as a share of time spent on work activities (Ellingrud et al., 2023).

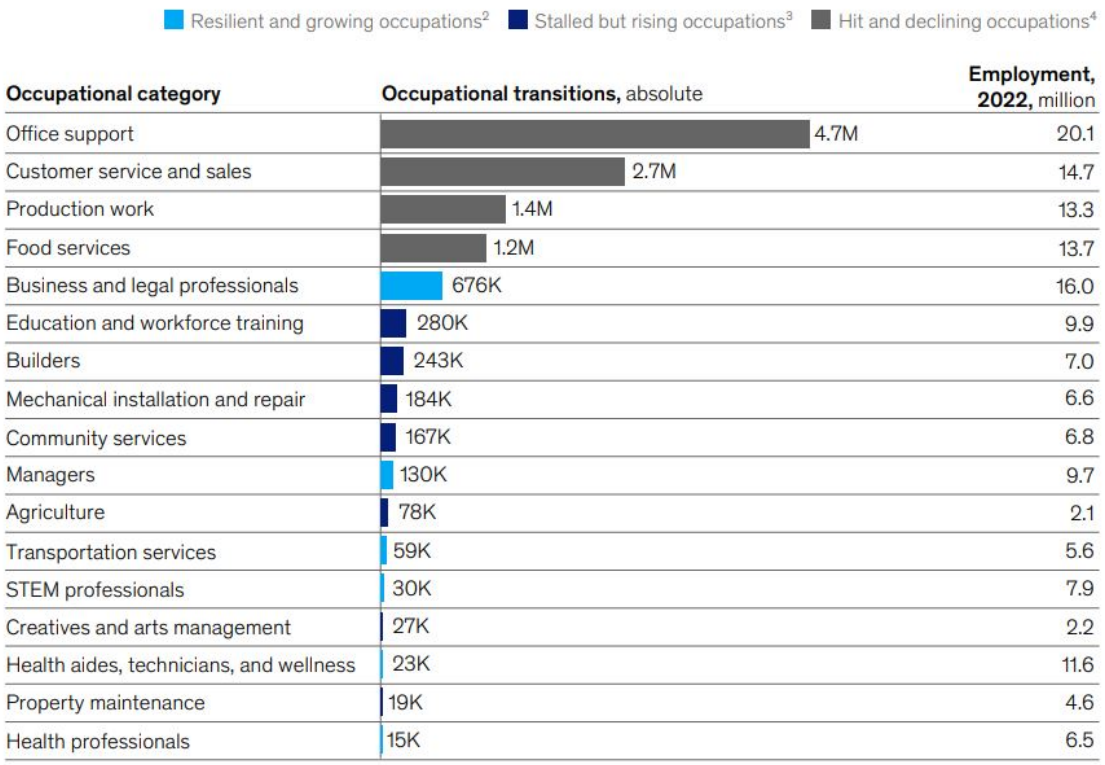
The implications of such acceleration can be drastic. The most straightforward implication, and concern of most common workers, is the job displacement. These concerns are not unfounded, and different studies (Ellingrud et al., 2023; Chen et al., 2022, Morandini et al., 2023) support them.

There are various occupations that are declining, not only due to automation but also due to other forces impacting employment (Ellingrud et al., 2023). The study of Ellingrud et al. (2023) provides the following insight into this decline caused by automation:

> "The biggest future job losses are likely to occur in office support, customer service, and food services. ... These jobs involve a high share of repetitive tasks, data collection, and elementary data processing, all activities that automated systems can handle efficiently. Our analysis also finds a modest decline in pro-

duction jobs despite an upswing in the overall US manufacturing sector, which is explained by the fact that the sector increasingly requires fewer traditional production jobs but more skilled technical and digital roles" (p. 42).

Figure 3.2 provides a graphic estimation of occupational transitions, with some alarming numbers of declining occupations:



Occupational category	Occupational transitions, absolute	Employment, 2022, million
Office support	4.7M	20.1
Customer service and sales	2.7M	14.7
Production work	1.4M	13.3
Food services	1.2M	13.7
Business and legal professionals	676K	16.0
Education and workforce training	280K	9.9
Builders	243K	7.0
Mechanical installation and repair	184K	6.6
Community services	167K	6.8
Managers	130K	9.7
Agriculture	78K	2.1
Transportation services	59K	5.6
STEM professionals	30K	7.9
Creatives and arts management	27K	2.2
Health aides, technicians, and wellness	23K	11.6
Property maintenance	19K	4.6
Health professionals	15K	6.5

Figure 3.2: Estimated number of occupational transitions by category, 2022–30 (Ellingrud et al., 2023).

As can also be observed in Figure 3.2, while some of the occupations such as office support and customer service are expected or already experiencing a decline, there are also occupations, that are more resilient and even expected to experience growth. The job that is expected to have the largest growth is expected to be healthcare, with an estimated 3.5 million more jobs by 2030 (Ellingrud et al., 2023). This estimation can also be supported by the study of Chen et al. (2022), pointing out the increased necessity of professionals in the healthcare sector with higher digital skills. Such professionals can provide a better quality

of patient treatment and more cost-effectiveness of work due to their higher efficiency in information and technology.

Other occupations that are expected to experience growth by 2030 are banking, insurance and pharmaceuticals, due to undergoing major digital transformations and the need for tech workers with advanced skills, led by STEM jobs with an estimated 23 percent increase (Ellingrud et al., 2023).

The creation of jobs and new occupations implies another nuance: the creation of new tasks. In the previous four decades, even though industrial robots, digital technologies, computer-controlled machines, and artificial intelligence replaced existing tasks and labor, emergence of new tasks ranging from engineering and programming functions to those performed by executive assistants, data administrators and analysts, and others took place (Acemoglu and Restrepo, 2018). It is also known, that humans have an advantage in new and more complex tasks compared to the AI, therefore strongly limiting complete automation of all the existing tasks (Bogen and Rieke, 2018; Acemoglu and Restrepo, 2018). Moreover, the introduction of automated technologies takes the form of the introduction of new, more complex versions of existing tasks, that must be taken over by human workers (Acemoglu and Restrepo, 2018).

Li (2022) in his paper has the following conclusion regarding job displacement:

> "...the onset of intelligent software systems, AI, and machine learning will not lead to mass unemployment. Instead, the likelihood is that many job functions will be downgraded or even disappear, while training, retraining, reskilling, and upskilling will be necessary to prepare today's students and workforce to be more creative to respond to the call of Industry 4.0" (p. 10).

To prepare the labor force for the new, more complex tasks and overall work with AI technologies, there is a strong necessity for systematic training, reskilling, education and overall digital literacy. The following section is going to explore this topic in more detail.

3.2 Digitalization, Upskilling and Reskilling

Digitalization is one of the major trends that refers to the changes associated with the application of digital technology in all aspects of human society and the conversion of analog data into digital form (Parviainen et al., 2017). However, the development of new technologies such as the Internet and machine learning and their application in different fields and occupations transformed digitalization into a more specific form, also known as industry 4.0.

Industry 4.0 is a term used for the description of virtual reality fusion systems based on traditional manufacturing that is transformed with different cyber-physical systems, the Internet, the Internet of Things (IoT), AI, machine learning, and more to create an intelligent production system (Li, 2022). The gradual integration of Industry 4.0 in recent years has had great implications on different levels of organizations, but especially on workers and skills that are expected to be in high demand in the years to come.

As mentioned in the previous section, automation, and therefore industry 4.0, is going to influence most existing occupations, especially ones with routine tasks that can be easily automated, such as office support, customer service, sales, and manufacturing. However, even the jobs that were previously thought to be impossible to automate are now going through similar changes. One such area is management: there are expectations, that AI and machine learning are going to provide decision support information to a company's board of directors by 2026 (Li, 2022). Considering the fact, that even decision-making processes on the highest level will be automated, it is only natural to expect a shift in core skills for most of the workers. The World Economic Forum in its Report in 2020 projected that half of all employees worldwide would need reskilling by 2025 (Li, 2022).

Morandini et al., (2023) in their study specifically focus on the impact of AI on upskilling and reskilling. They argue, that there has already been a clear automation trend in a variety of back-office processes, such as data entry, document management, customer service, and accounting, through the use of NLP and AI. The introduction of generative AI has further accelerated the automation trend, with generative AI being able to replace or mimic human transversal skills, such as communication, problem-solving, and conflict resolution. In some

cases, generative AI systems can even mimic human skills such as reasoning, problem-solving, and creativity. It becomes clear, that some skills will become obsolete if AI is integrated with other ICT skills, and from the ongoing trend of digitalization, it is to be expected. However, new sets of skills will have to emerge to replace those that were taken care of by the AI.

Multiple types of skills are expected to be changed in the years to come. Li, (2022) in his study has represented his findings regarding the top 10 most necessary skills for a future-ready workforce:

25/20/15*	in 2025	20/15*	in 2020	in 2015
1	Analytical thinking and innovation	1, 1	Complex problem solving	Complex problem solving
2	Active learning and learning strategies	2, 4	Critical thinking	Coordinating with others
3, 1, 1	Complex problem-solving	3, 10	Creativity	People management
4, 2, 4	Critical thinking and analysis	4, 3	People management	Critical thinking
5, 3, 10	Creativity, originality, and initiative	5, 2	Coordinating with others	Negotiation
6	Leadership and social influence	6	Emotional intelligence	Quality control
7	Technology use, monitoring, and control	7, 8	Judgment and decision making	Service orientation
8	Technology design and programming	8, 7	Service orientation	Judgment and decision making
9	Resilience, stress tolerance, and flexibility	9, 5	Negotiation	Active listening
10	Reasoning, problem-solving	10	Cognitive flexibility	Creativity

Figure 3.3: Top 10 skills on reskilling and upskilling future-ready work force (Li, 2022).

As seen in Figure 3.3, the table provides the comparison between the expected top 10 skills in 2020 and 2025. What is interesting in this comparison, is that seven out of 10 top skills listed under the column "in 2025" are not even listed under 2020 and 2015, while between 2015 and 2020 skill requirements overlap consistently. This tendency can be the result of increasing automation and especially the current AI development. This proposition can further be supported when looking at each of the separate skills: while 2020 and 2015 both have complex problem-solving as the most valuable skill, it dropped to third place in 2025. Overall it can be observed that almost any skill that can be automated or be somehow supported or completely replaced by AI has lost its relevancy in 2025, giving place to the more "uniquely human" skills that can be hardly automated, or not automated at all. At the same time analytical thinking and innovation, active learning and learning strategies are the new skills that are considered of the highest value in 2025.

Active learning being in the second position is another essential sign of the importance of the necessity of reskilling and upskilling the workers in organizations. Even though reskilling and upskilling both encompass learning new skills have some important key differences: upskilling usually refers to the process of acquiring new or improving existing skills relevant to the current field of work, meanwhile reskilling involves learning completely new skills outside one's current field (Morandini et al., 2023). Considering the current speed of development of the new technologies, organizations must be prepared for both of these learning processes and especially reskilling, if they want to stay competitive. Quoting Morandini et al. (2023):

> "Reskilling is an important process for organisations that introduces AI systems, as it can help employees adapt to the changes the technology brings. ... reskilling is crucial for companies looking to adopt AI as it helps employees develop the knowledge and skills they need to work effectively with the technology but in a new role" (p. 53).

Li (2022) further supports these statements:

> "... by focusing on scalable reskilling and upskilling, people would be fully equipped to participate in economic development, reducing inequality and leading to better social stability" (p. 10).

Another important contribution from Li's paper is made in the form of the proposition of a possible blueprint for reskilling and upskilling in organizations (see Figure 3.4):

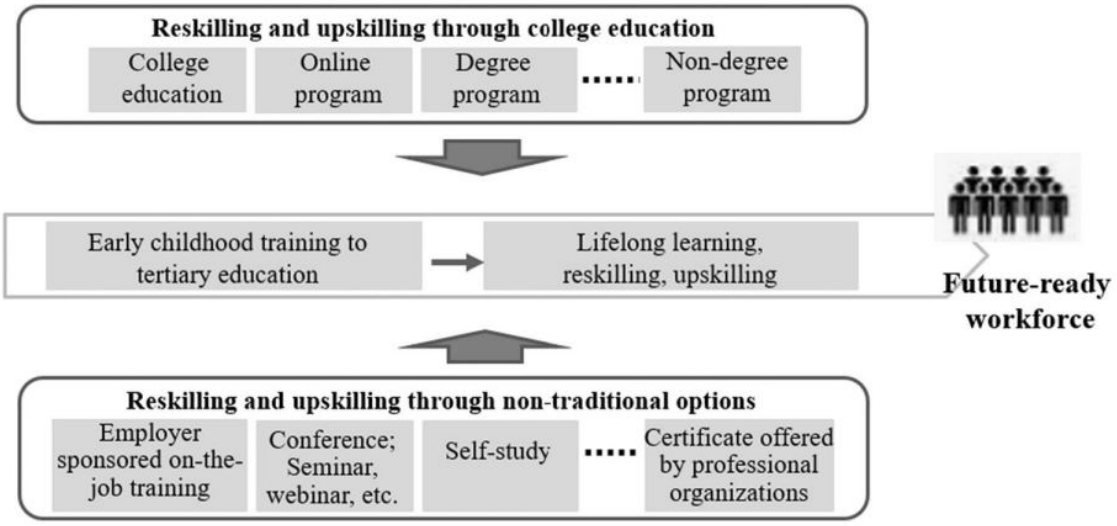


Figure 3.4: Blueprint of work-force upskilling and reskilling (Li, 2022).

The main purpose of the proposition consists in creating a framework that reinforces a life-long learning mechanism in the form of upskilling and reskilling through traditional options and other non-traditional options. In this framework childhood education, remains fundamental and mandatory for every citizen, however, it is further enhanced by traditional education in the form of colleges, degree and non-degree programs as well as online programs. On the other hand, after traditional learning, the worker will be offered non-traditional options such as employer-sponsored on-the-job training, seminars, self-study, and certificates (Li, 2022). The importance of collaboration between universities, government, and business organizations to become members of the education alliance to create a more effective ecosystem for life-long learning is further addressed as an important part of the process.

Certain obstacles must be addressed about reskilling and upskilling. Businesses will need to analyze how jobs and responsibilities could evolve, and workers will need to change their perspective to see new tools not as job destroyers but as work enhancers (Ellingrud et al., 2023). The lack of understanding of the impact of automation and digitalization on higher manager levels is another problem that should be addressed to create the necessary foundation for the reskilling of the workforce (Li, 2022). Accessibility, affordability and unwillingness of the older groups to upskill or reskill are also challenges that must be dealt

with to guarantee effective learning processes (Li, 2022; Morandini et al., 2023).

3.3 Economic Impact, Productivity and Efficiency

The beginning of the impact of AI on society can be compared with the impact of computers and computerization in general. Since the appearance of the calculating machines, computerization has had an impact on jobs that involve routine tasks, such as data entry, bookkeeping, among other tasks and jobs, reducing demand for workers performing "routine" tasks, at the same time increasing the productivity of workers who possess complementary skills, such as programming, data analysis, and research (Brynjolfsson et al., 2023; Ellingrud et al., 2023; Acemoglu and Restrepo, 2018). However, AI brings automation one step further, allowing the performance of non-routine tasks such as software coding, persuasive writing, and graphic design (Brynjolfsson et al., 2023).

The economic impact of AI and especially generative AI is yet hard to estimate due to its novelty. The most direct negative impact and consequence of its integration into organizations and society in general can be observed through job displacement (Ellingrud et al., 2023). For example, the recent stagnation of labor demand in the US can be explained by an acceleration of automation, particularly in manufacturing, and a deceleration in the creation of new tasks (Acemoglu and Restrepo, 2019). Acemoglu and Restrepo (2018) explain this trend as a result of financial decisions in favor of cost-saving:

> "If the elastic labor supply relationship results from rents (so that there is a wedge between the wage and the opportunity cost of labor), there is an important new distortion: because firms make automation decisions according to the wage rate, not the lower opportunity cost of labor, there is a natural bias toward excessive automation" (p. 39).

Despite a clear trend for excessive automation, there are also possible positive economic consequences that can originate from AI integration. Many research papers and other publications recognize the fact, that one of the main implications of the integration of

AI will be the automation of routine tasks (Brynjolfsson et al., 2023; Ellingrud et al., 2023; Acemoglu and Restrepo, 2018; Noy and Zhang, 2023), however, there may also be technological barriers to the automation of certain tasks and the creation of new tasks across industries, that would slow down the overall job displacement (Acemoglu and Restrepo, 2018).

Another problem that can be suggested through the insights of del Rio-Chanona et al. (2023), is that through the concentration of data and knowledge in the hands of private companies that own and control AI algorithms, a following concentration of wealth and power can be expected, leading to further economic inequality, especially if small businesses and individuals cannot afford access to these technologies. If such disruption will indeed take place, the economic impact will further increase the division between large and small enterprises, as well as between developed and developing regions (Ellingrud et al., 2023).

At the time of writing this work in 2024, humans still hold the advantage in new and more complex tasks. AI can not replace humans in solving these complex tasks, but it can greatly support them and enhance overall performance. Noy and Zhang (2023) have conducted an experiment with the implementation of ChatGPT, to find out the effects of the introduction of generative AI on grant writers and marketers.

In their experiment, 444 experienced, college-educated professionals were hired and were assigned each to complete two occupation-specific writing tasks. The idea was to create two groups, the treatment group, which was instructed to use ChatGPT between the first and the second task, and the control group, which had no access to it for any of the tasks. The results of their work were then checked and graded by independent professionals, who were not informed about the individual's group belonging. The results were very revealing: in the treatment group productivity and efficiency of the working process and the results very significantly improved. The time taken on the post-treatment task dropped by 37 percent relative to the control group and average evaluator grades in the treatment group increased by 0.45 standard deviations (see Figure 3.5, plots a) and b)). Not only that, but the treated group also took significantly less time to complete the second task, at the same time achieving a better score than the control group.

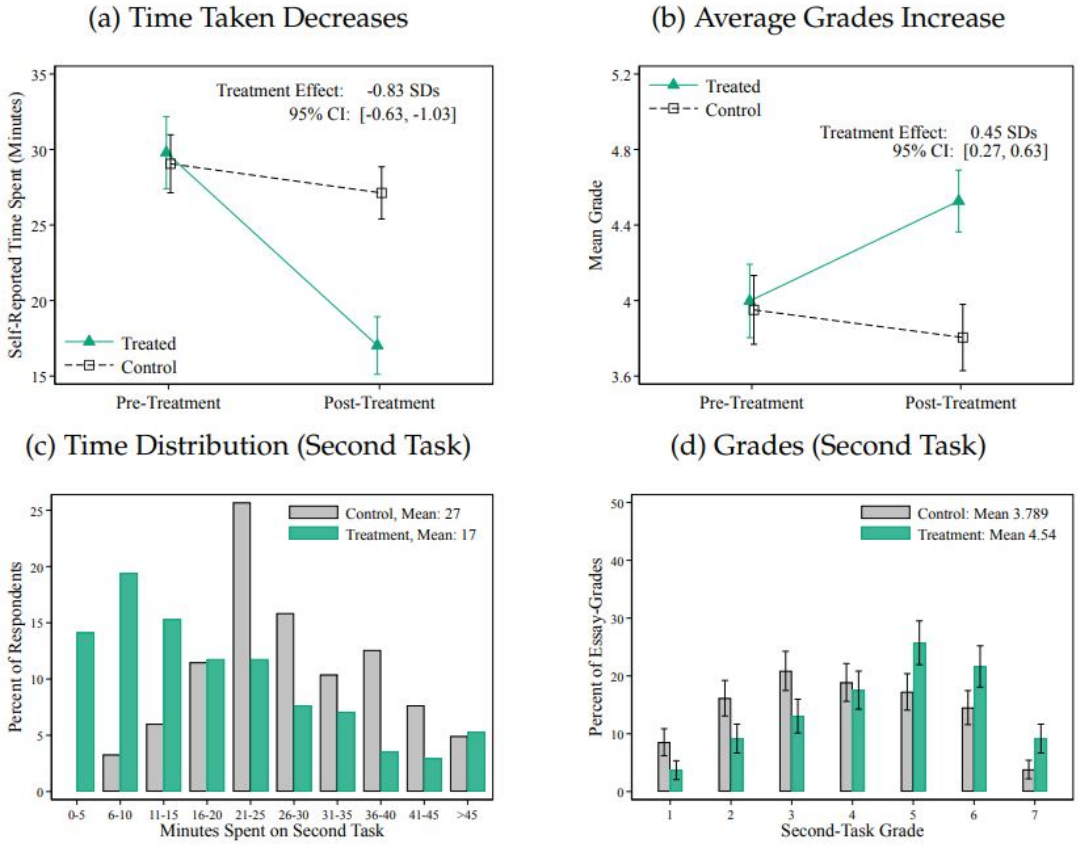


Figure 3.5: Effects of implementation of ChatGPT on productivity (Noy and Zhang, 2023).

An important detail about the performance of the treatment group is that 23 percent chose to completely replace their response with ChatGPT's output, while, 25 percent used ChatGPT to edit their original response, suggesting that the majority of the participants view ChatGPT as a way to improve output quality in addition to a convenient way to save time (Noy and Zhang, 2023).

This experiment proves the possibility of improvement of cognitive skills through the usage of generative AI. However, despite possible significant improvement in productivity and efficiency, the integration of AI technologies has two outcomes: either displace humans completely from certain occupations or complement existing human workers and increase their productivity (Brynjolfsson et al., 2023; Acemoglu Restrepo 2019; Noy and Zhang, 2023).

On the other hand, an optimistic possibility was also indicated by Ellingrud et al. (2023):

> "When machines take over dull or unpleasant tasks, people can be left with more interesting work that requires creativity, problem solving, and collaborating with others" (p. 54).

It is an optimistic outlook on the situation, with the hope that the integration of AI technologies could make workers happier by automating tedious or annoying components of the task or allowing them to concentrate on more interesting tasks, which would result in overall satisfaction (Noy and Zhang, 2023).

The following sections try to consolidate the theoretical background and statements with examples of the integration of AI technologies such as Copilot and ChatGPT in real life tasks and situations.

3.3.1 Generative AI in Sales

A very detailed and interesting perspective of the impact of AI on organizations and work is the case study conducted by Brynjolfsson et al. (2023), titled "Generative AI and Work". The case study analyses the introduction and adoption of a generative AI-based conversational assistant using data from 5,179 customer support agents.

The AI system introduced to the company in the study is based on a generative model that combines GPT with additional machine learning algorithms specifically fine-tuned for customer service interactions. This system is further trained on an extensive dataset of customer-agent conversations, labeled with various outcomes and characteristics, such as the success of the call resolution, the duration of the call, and the performance rating of the agent by the data firm.

Right from the beginning, the authors point out the positive impact of AI: access to the tool has increased the productivity of agents by almost 14 percent. As seen from Figure 3.6, panel A shows a decrease in the average handle time as soon as the AI tool was introduced in the company. Similarly, panel B shows the amount of chats, or calls that an agent could

handle per hour, increasing significantly.

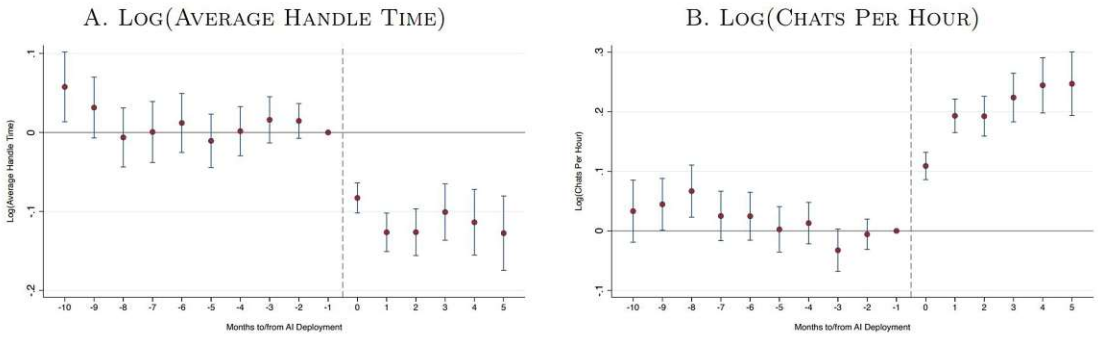


Figure 3.6: AI effect on customer service (Brynjolfsson et al., 2023).

On the other hand, another important finding was in relationship to the skill and tenure of an agent. It can be observed from Figure 3.7, how the agents of different skills, which are defined by their average call efficiency, resolution rate, and surveyed customer satisfaction in the quarter prior to the adoption of the AI system, change their performance after the introduction of the AI system. As the plot clearly shows, the agents with the lowest skill (Q1) have the biggest increase of 35 percent in resolutions per hour compared to their pre-AI adoption results. However, a clear trend can also be observed: the more skilled the agent is, the smaller the increase is, meaning that AI assistance does not lead to any productivity increase for the most skilled workers.

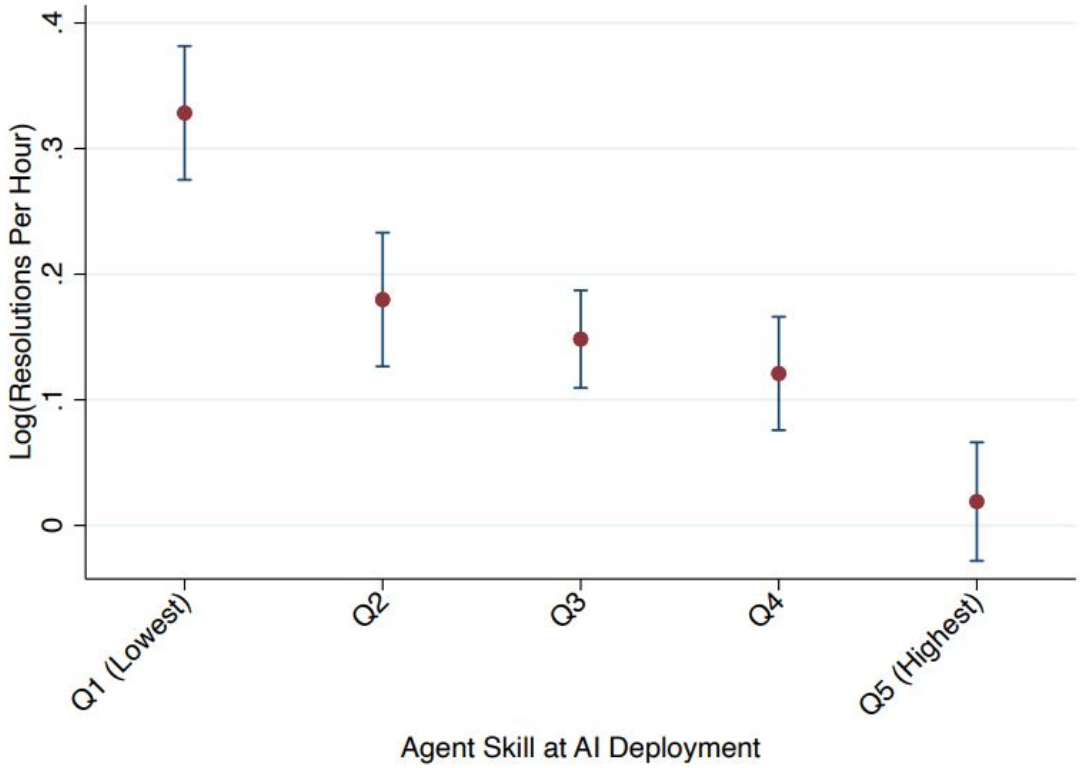


Figure 3.7: AI and agent skill (Brynjolfsson et al., 2023).

Apart from productivity, AI adoption has positively influenced less skilled agents on all other outcomes. Figure 3.8 shows that the average handle time has decreased (panel A), chats per hour (panel B), resolution rate (panel C) and even customer satisfaction (panel D) have increased. For the skilled workers, however, the results are mixed, but most importantly, showing negative tendencies in relationship to resolution rate (panel C) and customer satisfaction (panel D). These findings may suggest, that AI does not increase the efficiency of skilled workers, on the opposite, it may distract them from doing their jobs effectively.

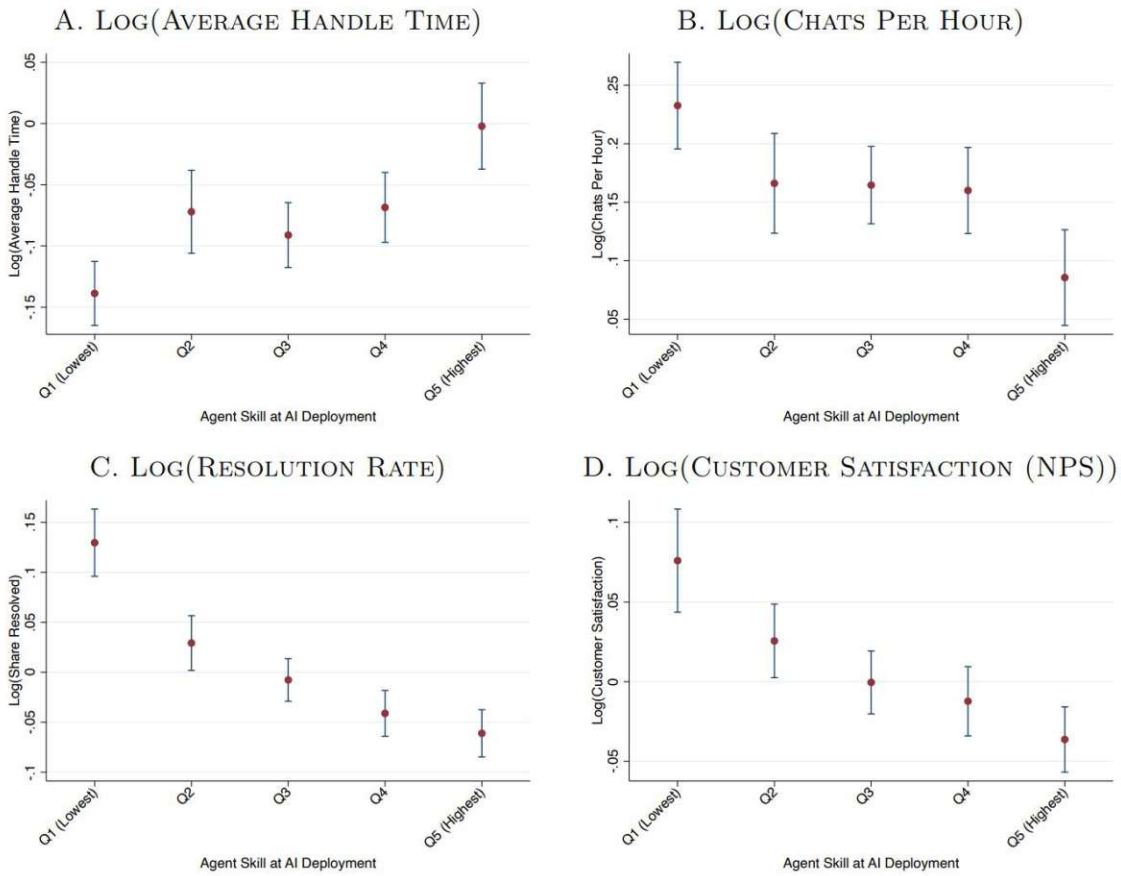


Figure 3.8: Additional findings of AI adoption (Brynjolfsson et al., 2023).

On a larger scale, the impact of AI can be seen through work experience, sentiment, and attrition. Through the provided data, it could be suggested, authors pointed out, that generative AI may improve agent's social skills and have a positive emotional impact on customers. Not only that but there is also a noticeable reduction in attrition, especially among newer workers, lowering the likelihood that a worker leaves in the current month. Finally, the adoption of AI was followed by an almost 25 percent decline in customer requests to speak to a manager, meaning the reduced number of escalation cases between the customer and the sales agent.

Based on the gathered data, Brynjolfsson et al. (2023) make four sets of findings:

1. AI assistance increases worker productivity through the above mentioned factors.
2. AI assistance disproportionately increases the performance of less skilled and less

experienced workers across all productivity measures considered.

3. High-skill workers may have less to gain from AI assistance, possibly due to training of the algorithm based on the data gathered from high-skill workers themselves. Lower-skilled agents, however, can improve by adhering to AI suggestions, basically incorporating behaviours of the higher-skilled workers.

4. The introduction of AI systems can impact the experience and organization of work.

This case study is one of the first of its kind, providing very valuable insights into the implications of generative AI adaptation in the customer service branch and already points out the potential and the advantages of this technology even in the earlier stages of its deployment.

3.3.2 Generative AI in Software Development

Another field that is expected to experience changes due to the appearance of generative AI is software development and IT in general. As mentioned in previous chapters, generative AI has the ability to understand and improve coding, thus creating new possibilities and challenges for workers in the field. Peng et al. (2023) have conducted a case study titled "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot" to find out precisely which effects generative AI has on this field.

For this case study, 95 professional programmers were hired and each of the participants was given the task of writing an HTTP server in JavaScript. The participants were divided into two groups: treatment and control groups. The treatment group used GitHub Copilot, an AI programmer powered by OpenAI's generative model, Codex, that suggests code and entire functions in real time based on context to complete the task, while the control group would not. The performance of the group was graded by two metrics: task success and task completion time, where task success was measured as the percentage of participants that successfully finished the task, and task completion time was the time from the start to finish of the task.

The results of this experiment can be seen in the Figure 3.9:

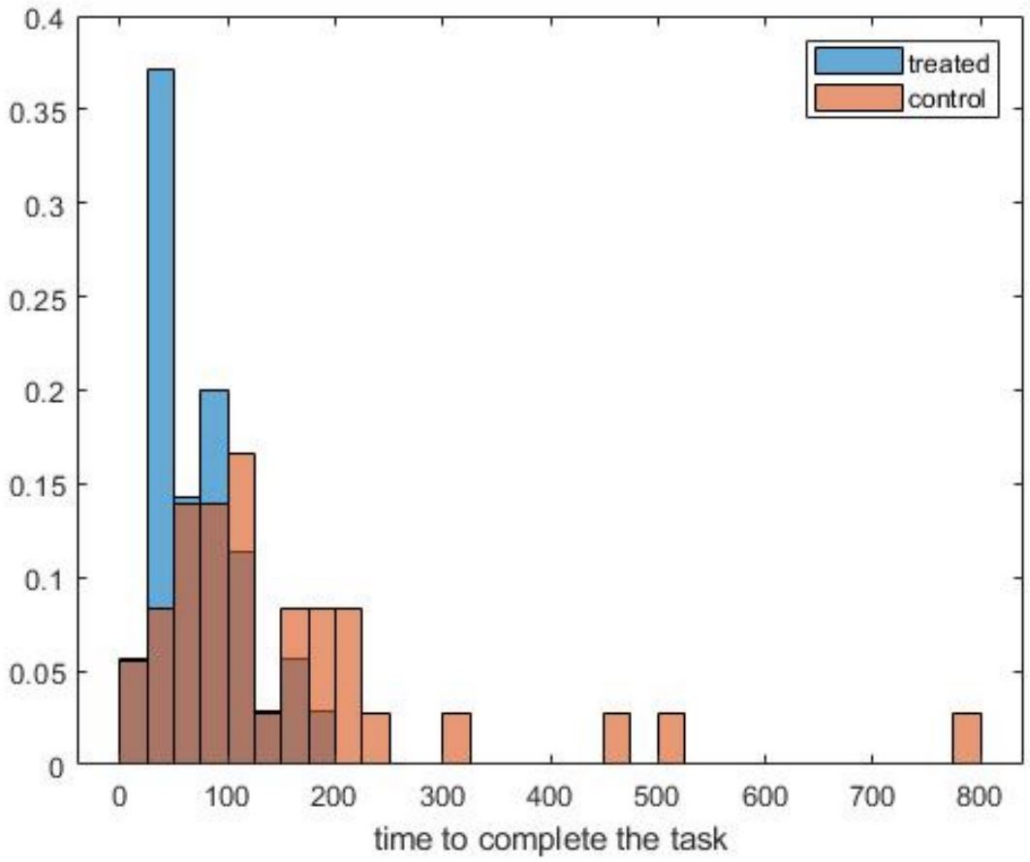


Figure 3.9: Task completion time (Peng et al., 2023).

The main finding and difference between both groups is the significant increase in productivity of the treated group: the average completion time from the treated group was 71.17 minutes and 160.89 minutes for the control groups, representing a 55,8 percent reduction in completion time. At the same time, the success rate of the treated group was, even though less significant, also higher than the one of the control group and measured at 7 percent. The correlation of these results with the personal attributes of participants has shown, that less experienced developers, as well as developers with heavy coding load (hours of coding per day), benefited more from Copilot usage. Another finding was the underestimation of the productivity of Copilot by both the control and the treated groups: both groups on average estimated a 35 percent increase in productivity, which is significantly less than the result of 55,8 percent.

The authors highlight in their case study the potential of generative AI to influence the

entire software development, pointing out the results of the experiment. Similarly to the case study of Brynjolfsson et al. (2023), less experienced workers are observed to benefit more from it. In conclusion, some ethical concerns are also being mentioned, such as security considerations due to code quality performance and the importance of upskilling, in order to stay competitive in the field.

3.3.3 AI in Radiology

A field that is also expected to be strongly influenced by AI in the coming years, is the field of medical care (Faynleyb, 2024; Ellingrud et al., 2023). A confirmation of this expectation can be found in the case study of Agarwal et al. (2023), titled "Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology". The authors point out, that AI has made significant advances in radiology, surpassing human decision-making in many cases, and such algorithms are already clinically deployed. Therefore, case studies can be conducted to find out the effects of AI on the field. In this case study, the efficiency of AI tools in the radiology field is being observed and compared to the human specialists. Despite not centered around a generative AI, this case study is fundamental in presenting problems and challenges regarding cognitive skills and trust that educated radiologists experience with the integration of this technology.

In this case study, professional radiologists were recruited through teleradiology companies to diagnose retrospective patient cases. To the provided chest X-ray image of a patient, either AI predictions in the form of probabilities or contextual information in the form of clinical history information, or both were added as support for the decision-making. This was made to estimate the treatment effects of both informational interventions on radiologists' prediction accuracy and the probability of making a correct decision. One of the effects studied, was the automation bias, presented in the form of a tendency to trust machine-provided predictions more than on one's conclusions.

One of the most important findings of the study is that AI has better prediction accuracy than approximately two-thirds of radiologists. Figure 3.10 shows the exact distribution of radiologists compared to the AI.

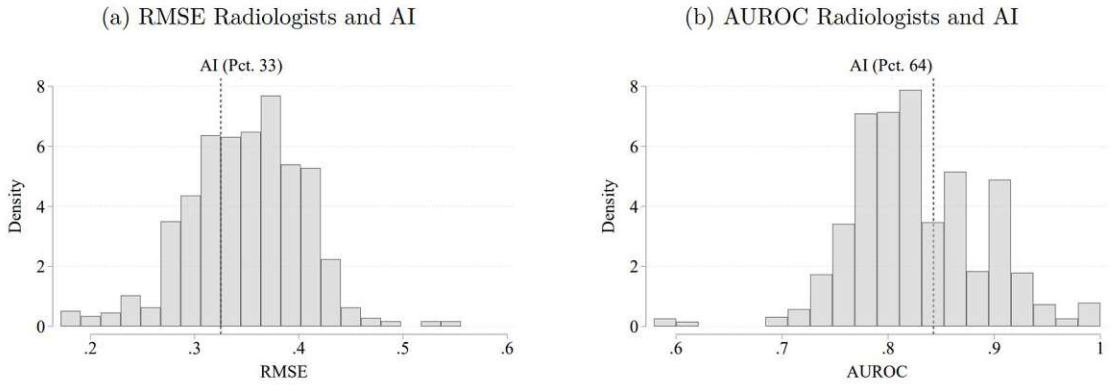


Figure 3.10: Accuracy of AI and radiologists (Agarwal et al., 2023).

Plot a) shows the root mean squared error (RMSE), which utilizes cardinal information about the AI prediction, where lower RMSE indicates a higher performance, whereas plot b) uses the receiver operating characteristic (ROC) curve, which "measures the trade-off between the false positive and the true positive rate of a classifier as the cutoff for classifying a case as positive or negative is varied" (Agarwal et al., 2023). A higher AUROC indicates a higher performance and vice-versa.

The most important findings, however, were the ones related to the diagnostic quality through the influence of AI. Quoting Agarwal et al. (2023):

> "...we observe that radiologists' reported probabilities move significantly towards the AI's predictions with AI assistance. Instead, the zero effect of AI assistance is driven by heterogeneous treatment effects – diagnostic quality increases when the AI is confident (e.g. the predicted probability is close to zero or one) but decreases when the AI is uncertain. In parallel, AI assistance improves diagnostic quality for patient cases in which our participants are uncertain, but decreases quality for patient cases in which our participants are certain. In contrast, providing clinical history does improve diagnostic quality suggesting that humans have additional valuable information that has not yet been incorporated into AI predictions" (p. 5).

As the authors themselves state, AI assistance has a mixed effect on the diagnostic quality,

depending on the uncertainty level of AI and the expert: with a confident AI diagnostic quality improves and decreases in the opposite case. This finding alone highlights the importance of a properly trained algorithm for achieving higher quality. What is also important, is how AI assistance improves the decision-making quality of doctors, when the doctors themselves are uncertain and decreases the decision-making quality of the confident ones, which undermines AI's usefulness for experienced professionals. On the other hand, providing clinical history has shown an overall positive effect on diagnostic quality.

Agarwal et al. (2023) conclude, that access to AI does not guarantee radiologists a higher performance, even though by itself AI outperforms two-thirds of them. Only with a confident AI tool should the AI assistance be considered. Moreover, access to AI also increased the average time to make a decision, which also works against AI-radiologist collaboration. Therefore, the final conclusion of the authors, is that radiologists should work *next to* as opposed to *with* AI.

Chapter 4

Conclusions

In this master thesis, several topics and issues were explored and analyzed from different perspectives. First of all, the theoretical background was presented. The concept and the basics of generative AI were introduced with a short explanation of LLMs and GPT, as well as current technologies already in use such as ChatGPT and Copilot. The second part of the theoretical background consisted of the ethical guidelines, where important issues such as bias and discrimination, privacy and data collection, transparency and accountability, automated decision-making, and autonomy were explored and analyzed, concluding with the introduction of the labour process theory that was further used for the analysis of the socio-economic impact.

Following the theoretical background, the socio-economic impact was thoroughly analyzed. Such subjects as job displacement, digitalization, upskilling, reskilling, productivity, and efficiency in relation to AI were researched and discussed. These findings were further supported and explored through examples of real-life adoption in organizations in different fields such as sales, IT and radiology.

With all the provided information it is now possible to answer the remaining research questions, defined at the beginning of this master thesis.

How does the integration of AI technologies change cognitive skills and overall job requirements?

The integration of AI technologies and especially generative AI, as multiple authors have found out, is already reshaping the way people are working and this change will only become more drastic in the coming years and further development of AI technologies. One of the most clear trends that can be observed and expected to strengthen in the near future is the automation of routine tasks, which has also long been a central concern of LPT. AI excels at handling repetitive, routine tasks such as data entry, basic analysis, and customer service inquiries, which, logically, decreases the demand for human workers to perform these tasks. On one hand, from LPT perspective, this replacement of repetitive tasks can be seen as a trend of deskilling, where workers' roles are diminished, allowing management to exert greater control and reduce labour costs. However, on practice appears a need for more creative and complex problem solving, which AI is incapable of replacing at the current stage of its development. Such skills as critical and analytical thinking and innovation, active learning, and, as just mentioned, complex problem solving will be extremely valuable in the future, where AI takes a stable and important part of the working culture.

It is also important to mention, how AI has given an advantage to less experienced workers. As seen from several case studies and experiments, the usage of generative AI has often compensated for the lack of experience and skill of various workers, significantly increasing their productivity and efficiency and, at the same time reducing the time needed for task completion. Even the workers with significantly less experience have shown performances matching or close to the expert level in respected fields. However, the benefits of AI tools for the experts and professionals with experience are questionable, to the point where the tool's efficiency is undermined. And it is important to mention, that despite giving a significant boost to the performance of the less experienced workers, AI does not replace the depth of understanding and judgment that comes with experience. Experienced workers still possess knowledge, insights and experience accumulated over time, which AI cannot replicate. Therefore, while AI may equalize less experienced workers and experts to some extent, it doesn't make less experienced workers inherently more skilled.

These results highlight the importance of digital literacy and how essential it will be for the workforce in general. As AI will replace more routine tasks, new skill set will be essential to be introduced through colleges and employers. Workers will have to learn how to use AI tools effectively, which requires continuous learning and adaptability, which, as LPT also suggests, can be achieved through upskilling and reskilling. These measures are the minimum that should be implemented in every organization in order to stay competitive and relevant in the age of digital technologies and AI.

Which socio-economic and ethical implications does the widespread adoption of generative AI technologies within organizations have?

It can be clearly observed, that AI has strongly influenced modern society and will only continue to expand its influence onto more fields. The ongoing trend of automation has caused several shifts that will define the future of work in this decade. The most explicit result of automation, that has concerned people even before this technology has appeared, is the job displacement. Generative AI can automate cognitive and even creative tasks previously considered secure from automation, such as content creation and interaction with people. Automation is expected to take over tasks accounting for up to 29,5 percent of total hours worked in all the sectors of the US economy by 2030. Such drastic change will have a serious impact on the job market, possibly causing layoffs in many sectors. On the other side, as history has shown with other technologies, the rise of generative AI could create new job opportunities and new roles that can't be foreseen just yet, equalizing the effects of job displacement. To minimize the impact and mitigate the worst-case scenario, reskilling and upskilling will have to become essential parts of the working culture.

The adoption of generative AI in organizations has shown an overall positive impact on the productivity and efficiency of workers, especially in sales and software development, which will motivate more organizations to integrate this technology. It must be stressed, once again, that to achieve significant improvement in the results, organizations must be

ready to invest in digital literacy and the upskilling of their workers and be prepared for further organizational changes that technological disruptions bring with them. It must also be considered, that not all fields or jobs may be fit for AI integration in its existing form, as was shown by the example of radiology experts, which were more efficient without AI support.

From an economic perspective, generative AI is also expected to have a strong impact. Companies that own and control generative AI technologies would dominate markets, leading to an increased concentration of wealth and power and to further aggravating economic inequality, especially if small businesses and individuals will not be able to afford access to these technologies. This concern is also addressed by LPT, highlighting the unfairness and disruption that the unequal distribution of technology causes.

The main issue with the adoption of AI in organizations at the moment of writing this master thesis is multiple ethical concerns that are being consciously or subconsciously overlooked. To this date, bias and discrimination can be found in most generative AI models, there are also issues with privacy and data collection causing concerns about cybersecurity and the overall lack of laws for the protection of sensitive information. The line of accountability is blurred due to lacking transparency of AI technology, which urges the development of a framework that would provide such transparency and ensure proper accountability. And finally, all of these issues are directly linked to the problem of autonomy and ethical decision-making. The blind trust in AI is an issue that must be addressed and worked on, since it has an extreme impact on decisions that affect people and such decisions should not be made uninformed or guided by a simple trust in the algorithm. Human autonomy should not be undermined by any technology, but rather be enhanced and supported, and AI has all the potential to be such a tool with a proper ethical design and development, that, for example, includes the HITL approach. Until all of the above-mentioned concerns are addressed and worked on, the integration of AI in organizations will be followed by the discussed issues that will negatively affect people and society in general. Therefore, the creation of regulations and governance, as well as public awareness and education are essential to mitigate these implications.

What recommendations can be given to the lawmakers and organizations that begin to implement AI?

All the ethical and socio-economic concerns require close attention to mitigate the negative consequences that were previously mentioned. Regulations and governance are a good foundation for the ethical design of AI. The lawmakers can be advised the development of comprehensive AI regulations, that establish clear ethical guidelines focusing on such topics as bias and discrimination, privacy and data collection, as well as transparency and accountability. Bias and discrimination should be addressed and fought against via unbiased training data and thorough testing of AI algorithm, that would prevent their manifestation at the root. AI systems should also be developed to comply with stricter data privacy laws, protecting personal data and preventing its misuse. Growing concerns of potential cybersecurity issues should also be addressed to prevent future misuse of generative AI for security breaches and cyber attacks in general. Transparency requirements in AI decision-making processes should also be implemented, requiring companies to disclose how AI systems make decisions and the data they rely on. Human autonomy and ethical decision-making should be at the center of these regulations, allowing the development of AI systems that will enhance and support, rather than undermine them. Overall, more support for AI research and development in the form of investments and international collaborations consisting of different stakeholder groups should be created.

For organizations that are beginning to integrate AI tools, it is essential, that everyone will be informed and educated about the technology, its benefits and, most importantly, its potential risks. Therefore, AI education and training are the minimum, that organizations should provide their employees to maximize the positive effects and mitigate the risks and disadvantages. The implementation itself should be gradual and careful, for example, with small projects, allowing the organization to learn and adapt before further deployment. It is also strongly advised, that the integration is happening under constant human oversight, ensuring a proper implementation, that allows identification of issues or biases that may emerge over time and would enable iterative improvements of the AI system.

Afterword

AI is a technology that exceeds humans in almost every way. It has proved itself by enhancing human productivity and efficiency or in some cases completely replacing humans in certain tasks. This master thesis explored how it already has affected and will possibly affect society in the future years, and it seems, that AI has the potential to become an inseparable part of our lives, just like smartphones and the internet, integrated in a way that can not even be foreseen just yet. People will trust it and become dependent on it. Therefore, it is ever more needed to explore and analyze all possibilities of the development of this technology and address the risks before the damage has been done.

We need to accept, that despite being a powerful tool of the future, it is still a flawed and imperfect technology, that will have to undergo various changes, improvements and upgrades before it will be ethically reliable. We, as engineers, developers, and ethicists, are responsible for AI's further growth and refinement and can ensure, that it will become what we envision it to be and we should do our best to "teach" it the best of what humanity has to offer. After all, quoting the author from The Atlantic article (What Isaac Asimov Can Teach Us About AI, 2023), Isaac Asimov came to this realization first:

> "What AI learns, actually, is to be a mirror - to be more like us, in our messiness, our fallibility, our emotions, our humanity."

Acknowledgement

In this section, I would like to thank everyone who has helped me to complete this master thesis.

My special thanks are addressed to Univ.Prof.in Mag.a rer.soc.oec. Sabine Theresia Köszegi, who provided me the opportunity to work on this master's thesis. I'm grateful to her for all the support and encouragement that was needed to complete this thesis. She was always available for questions and helped to solve problems and complications that appeared during the process.

I would also like to thank all of my family members, friends and fellow students who supported me all the way through this journey. Only thanks to you the completion of this study was possible.

Bibliography

Acemoglu, D. and Restrepo, P. (2018). *The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment.* American Economic Review, 108(6), pp.1488–1542.

Acemoglu, D., and Restrepo, P. (2019). *Automation and new tasks: How technology displaces and reinstates labor.* Journal of economic perspectives, 33(2), 3-30.

Acemoglu, D., Autor, D., Hazell, J. and Restrepo, P. (2022). *Artificial Intelligence and Jobs: Evidence from Online Vacancies.* Journal of Labor Economics, 40(S1), pp.S293–S340. doi:https://doi.org/10.1086/718327.

Adetayo, A. J., Aborisade, M. O., and Sanni, B. A. (2024). *Microsoft Copilot and Anthropic Claude AI in education and library service.* Library Hi Tech News.

Agarwal, N., Moehring, A., Rajpurkar, P., and Salz, T. (2023). *Combining human expertise with artificial intelligence: Experimental evidence from radiology* (No. w31422). National Bureau of Economic Research.

Akinrinola, O., Okoye, C. C., Ofodile, O. C., and Ugochukwu, C. E. (2024). *Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability.* GSC Advanced Research and Reviews, 18(3), 050-058.

Allhutter, D., Cech, F., Fischer, F., Grill, G., and Mager, A. (2020). *Algorithmic profiling of job seekers in Austria: How austerity politics are made effective.* frontiers in Big Data, 3, 502780.

Araujo, T., Helberger, N., Kruikemeier, S., and De Vreese, C. H. (2020). *In AI we trust?*

Perceptions about automated decision-making by artificial intelligence. AI and society, 35, 611-623.

Baidoo-Anu, D., and Ansah, L. O. (2023). *Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning.* Journal of AI, 7(1), 52-62.

Beauchamp, T. (2008). *The principle of beneficence in applied ethics.*

Bogen, M. and Rieke, A. (2018). *Help wanted: an examination of hiring algorithms, equity, and bias.*

Brynjolfsson, E., Li, D. and Raymond, L. R. (2023). *Generative AI at work.* (No. w31161) National Bureau of Economic Research.

Chen, N., Li, Z., and Tang, B. (2022). *Can digital skill protect against job displacement risk caused by artificial intelligence? Empirical evidence from 701 detailed occupations.* PLoS One, 17(11), e0277280.

Davenport, T. H., and Harris, J. G. (2005). *Automated decision making comes of age.* MIT Sloan Management Review, 46(4), 83.

del Rio-Chanona, M., Laurentsyeva, N., and Wachs, J. (2023). *Are large language models a threat to digital public goods? evidence from activity on stack overflow.* arXiv preprint arXiv:2307.07367.

Ellingrud, K., Sanghvi, S., Madgavkar, A., Dandona, G. S., Chui, M., White, O and Hasebe, P. (2023). *Generative AI and the future of work in America.*

Faynleyb, I. (2024) *The Impact of Automated Decisions on Role Expectations, Autonomy and Responsibility at Work.* TU Wien, Institute of Management Science

Ferrara, E. (2023). *Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies.* Sci, 6(1), 3.

Ferrer, X., Van Nuenen, T., Such, J. M., Cot´e, M., and Criado, N. (2021). *Bias and discrimination in AI: a cross-disciplinary perspective.* IEEE Technology and Society Magazine, 40(2), 72-80.

Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. (2024). *Generative ai.* Business and Information Systems Engineering, 66(1), 111-126.

Gandini, A. (2019). *Labour process theory and the gig economy.* Human relations, 72(6), 1039-1056.

Gupta, M., Akiri, C., Aryal, K., Parker, E., and Praharaj, L. (2023). *From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy.* IEEE Access.

Harris, J. G., and Davenport, T. H. (2005). *Automated decision making comes of age.* MIT Sloan Management Review, 46(4), 2-10.)

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., and Awadalla, H. H. (2023). *How good are gpt models at machine translation? a comprehensive evaluation.* arXiv preprint arXiv:2302.09210.

Kim, B., Park, J., and Suh, J. (2020). *Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information.* Decision Support Systems, 134, 113302.

Knights, D., and Willmott, H. (Eds.). (2016). *Labour process theory.* Springer.

Laitinen, A., and Sahlgren, O. (2021). *AI systems and respect for human autonomy.* Frontiers in artificial intelligence, 4, 705164.

Larsson, S., and Heintz, F. (2020). *Transparency in artificial intelligence.* Internet Policy Review, 9(2).

Li, L. (2022). *Reskilling and upskilling the future-ready workforce for industry 4.0 and beyond.* Information Systems Frontiers, 1-16.

Marx, K. (1976). *Capital: A Critique of Political Economy volume one.*. Penguin Books.

Microsoft (2024). *Microsoft Copilot | Microsoft AI.* [online] www.microsoft.com. Available at: https://www.microsoft.com/en-us/microsoft-copilot. (Accessed 5 May, 2024).

Morandini, S., Fraboni, F., De Angelis, M., Puzzo, G., Giusino, D., and Pietrantoni, L. (2023). *The impact of artificial intelligence on workers' skills: Upskilling and reskilling in*

organisations. Informing Science: The International Journal of an Emerging Transdiscipline, 26, 39-68.

Noy, S., and Zhang, W. (2023). *Experimental evidence on the productivity effects of generative artificial intelligence.* Science, 381(6654), 187-192.

OpenAI (2023). *GPT-4 Technical Report.* arXiv (Cornell University). doi:https://doi.org/10.48550/arxiv.2303.08774.

OpenAI (2024). *OpenAI.* [online] OpenAI. Available at: https://openai.com/ (Accessed 28 Apr., 2024).

Parviainen, P., Tihinen, M., Kääriäinen, J., and Teppola, S. (2017). *Tackling the digitalization challenge: how to benefit from digitalization in practice.* International journal of information systems and project management, 5(1), 63-77.

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). *The impact of ai on developer productivity: Evidence from github copilot.* arXiv preprint arXiv:2302.06590.

Rubel, A., Castro, C., and Pham, A. (2020). *Algorithms, Agency, and Respect for Persons.* Soc. Theor. Pract. 46 (3), 547–572. doi:10.5840/soctheorpract202062497

UNI Global Union. *The Future World of Work. Top 10 Principles for Ethical Artificial Intelligence*; UNI Global Union: Nyon, Switzerland, 2017.

Varona, D., and Suárez, J. L. (2022). *Discrimination, bias, fairness, and trustworthy AI.* Applied Sciences, 12(12), 5826.

What Isaac Asimov Can Teach Us About AI. (2023). theatlantic.com. Available at: https://www.theatlantic.com/books/archive/2023/03/ai-robot-novels-isaac-asimov-microsoft-chatbot/673265/. (Accessed 6 Aug., 2024).

Xiao, Y., and Watson, M. (2019). *Guidance on conducting a systematic literature review.* Journal of planning education and research, 39(1), 93-112.

List of Figures