POLITECNICO DI TORINO

Master of Science in
Engineering and Management

Master Thesis

"The influence of Artificial intelligence on productivity in Software development"

Tutor:
Prof. Anna D'Ambrosio

Candidate:
Daria Glushkova

2023

Content

Introduction

Over the past few years, artificial intelligence (AI) has made remarkable strides in various domains, transforming the way we live, work, and interact. The rapid advancement of AI technologies has significantly impacted numerous industries Czarnitzki et al (2022). For instance, in healthcare, AI algorithms can now analyze complex medical data and assist healthcare professionals in making more accurate and timely decisions, ultimately enhancing patient outcomes. In the financial sector, the use of AI-powered chatbots and virtual assistants has significantly improved customer experience and streamlined various financial processes.
The rapid advancement of artificial intelligence (AI) technologies has significantly impacted various industries, with software development being no exception.
The increasing complexity and scale of software development projects necessitate efficient approaches to ensure timely delivery and maintain competitiveness in the market. As AI-powered tools and technologies continue to evolve and gain prominence, companies continue assessing the possibilities to use them in order to capitalize on these advancements and drive innovation.

As AI systems continue to evolve, they offer potential solutions to enhance productivity by automating repetitive tasks, assisting in decision-making, and providing intelligent insights. In the end of 2022 OpenAi released ChatGPT – AI-driven language model which immediately captured the attention of the world, demonstrating remarkable capabilities in generating human-like text, providing valuable assistance in numerous tasks.
Giving that it is a recent development, there is a pressing need to examine its impact on productivity in software development. Understanding how AI-driven language models like ChatGPT can be integrated into software development workflows and the benefits they may offer can help organizations optimize their processes, enhance efficiency, and ultimately, achieve greater success in the competitive landscape.
This study is timely and relevant as it seeks to address the knowledge gap by examining the effects of ChatGPT on productivity in software development through a survey-based analysis. By collecting data from professionals in different roles and organizations, this research aims to provide valuable insights into how ChatGPT impacts various stages of the software development process. The findings of this study will contribute to the ongoing discourse on AI's role in software development, informing practitioners and researchers alike on the potential benefits and challenges of implementing AI technologies in the industry.

Research Problem

Although artificial intelligence technologies have demonstrated immense potential in various industries, their influence on productivity in software development, specifically when utilizing AI-driven language models like ChatGPT, remains underexplored. As claimed by Raj and Seamans (2018) and Czarnitzki et al (2022) until now there has not been a database available at the firm level that allows a rigorous study of the role of AI on productivity. With the growing interest in ChatGPT and its potential applications in software development, there is a pressing need to examine its impact on productivity to understand the benefits and challenges that its implementation may pose. Identifying the extent of

ChatGPT's influence on software development productivity can provide valuable insights for organizations seeking to optimize their processes and drive innovation in the competitive landscape.

Research Question

The primary research question for this study is: how does the implementation of artificial intelligence, specifically ChatGPT, affect productivity on different stages of software development processes?

Research aim

The aim of this research is to investigate the impact of implementing artificial intelligence, particularly ChatGPT, on productivity during various stages of software development processes.

Research goals

To achieve the aim of this research, the following goals have been identified:

1. Explore the evolution and capabilities of AI, including ChatGPT's applications in various industries.
2. Analyse productivity in software development and review existing studies on AI's impact on productivity.
3. Assess how ChatGPT influences productivity in different stages of software development.
4. Investigate correlations between ChatGPT's influence and respondent characteristics (e.g., grade, experience, company size).
5. Conduct ANOVA tests to identify differences in ChatGPT's influence on productivity among different respondent groups.
6. Examine which stages benefit most from ChatGPT during the software development process.

Research Methodology:

1. Research Design: The study utilized a mixed approach, however, setting quantitative as the main one to gain a data-proven comprehensive understanding of ChatGPT's influence on productivity in software development.
2. Participants and Sampling: The study utilized a purposive sampling strategy to target software development professionals with experience in using ChatGPT. The goal was to find people involved into different software development stages, from various industries and countries, with various background. The survey was distributed through several channels, including LinkedIn, Telegram, and WhatsApp.
3. Data Collection: The primary data collection method is an online survey, consisting of both closed-ended and open-ended questions. The data collection period spanned from March to July 2023, during which over 3000 messages were sent, and 150 responses were obtained.
4. Data Analysis: The data obtained from the survey were analyzed using statistical software JASP. Descriptive statistics, correlation analysis, ANOVA were made to identify patterns and trends in the data. Qualitative data from open-ended questions was used to make the picture more comprehensive and to clarify some of the insights received from the quantitative results.

Relevance

The relevance of this research lies in its examination of the influence of AI-driven language models, specifically ChatGPT, on productivity in the software development process. As AI technologies continue to advance and permeate various industries, understanding their potential impact on productivity is crucial for organizations seeking to optimize their operations and maintain a competitive edge. Software development, in particular, is a rapidly evolving field that demands constant innovation and efficiency improvements. Investigating the role of AI in enhancing productivity within this context is highly relevant for both practitioners and researchers.

Novelty

This research is new in its focus on ChatGPT, a relatively new AI language model, and its potential influence on productivity in software development. While there is a growing body of literature on AI and its applications in various industries, the specific impact of ChatGPT on software development productivity has not been extensively explored. By conducting a survey-based analysis involving software development professionals and examining their experiences and perceptions of ChatGPT's implementation, this study aims to fill this gap in the literature.

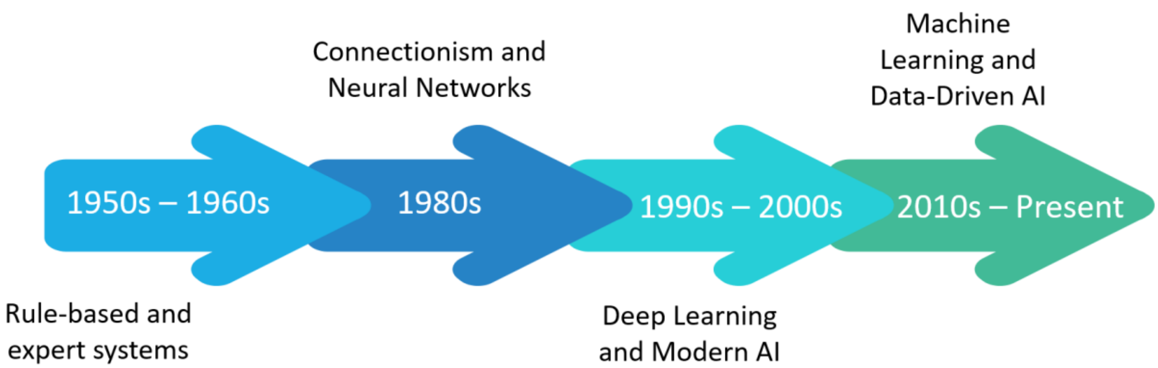
Chapter 1. Theoretical background
1.1. Artificial Intelligence: evolution, capabilities and applications

In this section, we will dive into the world of artificial intelligence (AI), exploring its historical development and the practical applications it offers across various industries. Understanding the evolution of AI is crucial to understand the current state of technology and the possibilities it brings. We will discuss the different types of AI, with a particular focus on AI language models, examining their capabilities and the ways they have revolutionized various sectors

1.1.1 Evolution of AI and its diverse applications

The development of Artificial Intelligence (AI) can be traced back to the mid-20th century, with the term "artificial intelligence" first coined by John McCarthy in 1955 (McCarthy et al., 1955). AI has come a long way since its inception, with significant milestones marking its evolution. This section highlights the key developments in AI history and provides an overview of its applications across various industries.

Figure 1. Main stages of evolution of AI



Source: made by author

Early AI: Rule-based systems and expert systems (1950s and 1960s)

- Characteristic: These systems were designed to mimic human decision-making by codifying expert knowledge into explicit rules.
- How it's different: Unlike later AI forms that learned from data, rule-based systems followed predetermined logic based on human expertise.
- Example: The General Problem Solver (GPS) by Newell and Simon (1961) attempted to simulate human problem-solving methods using these rules.
- Breakthrough principle: Rule-based systems relied on a clear and predefined set of rules, often handcrafted by experts. When a certain condition was met, the system would execute a corresponding action.

Connectionism and Neural Networks (1980s)

- Characteristic: This era saw a shift from rule-based logic to models that tried to emulate the way human brain's neurons interacted, known as neural networks.
- How it's different: Instead of relying on preset rules, these systems began to adapt and change their behavior based on input, thus learning over time.
- Key advancement: The backpropagation algorithm by Rumelhart, Hinton, and Williams (1986) transformed neural network training, setting the stage for deep learning.
- Breakthrough principle: Backpropagation adjusts the weights of connections in the neural network by calculating the gradient of the loss function. This iterative optimization refines the model's predictions over time.

Machine Learning and Data-Driven AI (Late 20th and early 21st centuries)

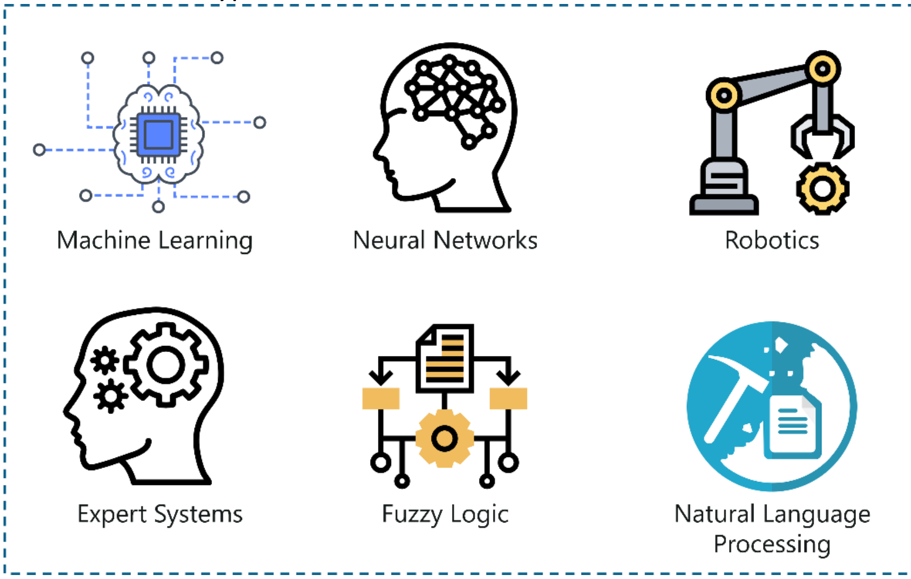
- Characteristic: Machine learning focuses on enabling systems to learn and make decisions from vast amounts of data without being explicitly programmed.
- How it's different: While neural networks were an early form of learning systems, machine learning brought in more diverse techniques that made AI more versatile.
- Notable techniques: Decision trees, support vector machines, and clustering algorithms (Bishop, 2006).

Deep Learning and Modern AI (21st century)

- Characteristic: Deep learning, a subfield of machine learning, uses layered neural networks (often very deep) to analyze various forms of data.
- How it's different: Deep learning models can autonomously extract features from raw data, making them particularly effective for complex tasks like image and speech recognition.
- Breakthrough: The development of convolutional neural networks (CNN) by LeCun et al. (1998).
- Breakthrough principle: CNNs are specialized for processing structured grid data, like images. They use convolutional layers to scan input data (like an image) with filters, capturing local patterns, and then pooling layers to reduce dimensionality. This structure enables CNNs to recognize hierarchical patterns efficiently.

Currently, AI can be classified by types according to Figure 2.

Figure 2. Domains of AI Types



Source: Edureka.com

Machine Learning (ML):

- Origin in Evolution: Stemming from the era of Machine Learning and Data-Driven AI, ML focuses on training models on vast datasets. While neural networks began the journey into data-driven insights, ML expanded the toolbox with various algorithmic techniques.
- Description: A subset of AI that enables computers to refine their functions over time by learning from data. It identifies patterns and makes predictions or decisions based on these patterns.
- Applications: Online recommendation systems in platforms like Netflix or Amazon use ML to analyze user preferences and suggest content or products.

Neural Networks:

- Origin in Evolution: Neural Networks find their roots in the era of Connectionism and Neural Networks when AI began to mirror the interconnectivity of the human brain.
- Description: These consist of layers of artificial neurons that process and learn from data, emphasizing pattern recognition and data classification.
- Applications: Neural networks underpin many modern AI applications, like face recognition in social media platforms where the system learns to identify features of a face and tag users.

Robotics:

- Origin in Evolution: Robotics, as a concept, has been influenced by all stages of AI evolution. Robots often employ a combination of rule-based logic (from the Early AI era) and machine learning.
- Description: This domain emphasizes the design and application of robots, which can operate tasks with varying degrees of autonomy.
- Applications: Robotic vacuum cleaners in homes use sensors and apply rule-based logic to navigate and clean rooms (Jurafsky, D. et al, 2019).

Expert Systems:

- Origin in Evolution: Direct descendants of the Early AI phase, expert systems encapsulate rule-based logic and human expert knowledge.
- Description: AI programs that simulate human expert decision-making in specific domains using pre-defined rules.
- Applications: Medical diagnosis systems, where symptoms inputted can lead to potential disease identifications based on medical knowledge.

Fuzzy Logic:

- Origin in Evolution: Fuzzy Logic can be seen as an advanced iteration of rule-based systems from the Early AI phase, providing more nuanced decision-making.
- Description: It offers reasoning solutions in situations with uncertain or incomplete data by working with degrees of truth rather than absolute true/false logic (Siciliano B. et al., 2016).
- Applications: Climate control systems in cars, where the system adjusts temperature based on various factors, not just the set temperature.

Natural Language Processing (NLP):

- Origin in Evolution: NLP has been highly influenced by the Deep Learning and Modern AI phase, leveraging advanced neural networks for language tasks.
- Description: Focuses on enabling computers to interact, understand, and generate human language.
- Applications: Chatbots on customer service websites use NLP to understand user queries and respond in a coherent manner.

AI has found applications in a multitude of industries, including healthcare, finance, transportation, and manufacturing.

Table 1: AI Applications in Various Industries and Examples

Industry	Application	Example System/Company
Healthcare	AI-powered diagnosis	Analyze medical images and patient data to detect diseases and suggest treatment plans. E.g., IBM Watson Health
	Drug discovery	Helps identify potential drug candidates, reducing the time and cost of drug development. E.g., Atomwise
Finance	Fraud detection	Identify fraudulent transactions and patterns by analyzing large datasets. E.g., Mastercard's Decision Intelligence
	Personalized financial advice	AI-powered chatbots and virtual assistants provide tailored financial guidance. E.g., Bank of America's Erica
Transportation	Autonomous vehicles	Enable self-driving cars to navigate based on real-time data. E.g., Tesla Autopilot
	Traffic management	AI algorithms analyze traffic patterns to reduce congestion. E.g., IBM's Intelligent Transportation

Manufacturing	Predictive maintenance	Predict equipment failures and optimize maintenance schedules. E.g., General Electric's Predix
	Supply chain optimization	Analyze data for inventory management and logistics. E.g., IBM Watson Supply Chain
Agriculture	Precision agriculture	AI algorithms optimize crop growth and monitor soil conditions. E.g., John Deere
	Crop and livestock management	Identify potential diseases, pests, and monitor livestock health. E.g., Cainthus' facial recognition for livestock
Retail	Personalized recommendations	Analyze customer behavior to suggest products. E.g., Amazon
	Inventory management	Predict customer demand and optimize stock levels. E.g., Walmart
Human resources	Talent acquisition	Streamline the hiring process with AI-driven tools. E.g., Infinity Innovators
	Employee performance management	Analyze employee data for insights into performance. E.g., Humu's nudges
Education	Adaptive learning platforms	Personalize content based on individual needs. E.g., DreamBox
	Automated grading and assessment	Evaluate and grade student work with AI. E.g., Turnitin
Marketing and Advertising	Targeted advertising	Analyze user data to deliver personalized ads. E.g., Google Ads
	Content creation	Tools generate human-like text for various uses. E.g., OpenAI's GPT-3
Entertainment and Media	Content recommendation	AI suggests content based on user habits. E.g., Netflix
	Video game development	AI-powered procedural generation for unique gaming experiences. E.g., No Man's Sky
Energy and Utilities	Smart grid management	Analyze and optimize energy distribution. E.g., GE's Grid Solutions
	Renewable energy forecasting	Predict renewable energy generation. E.g., DeepMind for Google's wind farms
Legal and Regulatory Compliance	Legal document analysis	Assist in reviewing legal documents. E.g., ROSS Intelligence
	Regulatory compliance check	Help organizations stay compliant by analyzing potential risks. E.g., IBM Watson Regulatory Compliance

Source: made by author

The table presented outlines a comprehensive overview of how AI is revolutionizing various industries. Each industry leverages AI in distinct ways, adapting its capabilities to address specific challenges and optimize operations.

Healthcare, for instance, is using the power of AI for diagnostic purposes. IBM Watson Health has been one of the leaders, using AI to parse medical images and patient data to detect diseases. Similarly, companies like Atomwise employ AI in drug discovery.

In the financial sector, AI's potential in fraud detection is evident with tools like Mastercard's Decision Intelligence. It efficiently identifies patterns of activities that look like fraud. Furthermore, AI-powered chatbots like Bank of America's Erica are redefining the banking experience, offering personalized financial advice to clients.

Transportation has been radically transformed with the advent of autonomous vehicles. Tesla's Autopilot, for example, navigates the complexities of real-time driving using advanced AI algorithms. Meanwhile, traffic management systems, such as IBM's Intelligent Transportation solution, are reducing congestion by optimizing traffic flow based on data patterns.

In manufacturing, predictive maintenance and supply chain optimization, as seen with General Electric's Predix and IBM Watson Supply Chain respectively, are ensuring operations run smoother and more efficiently.

The applications of AI in industries like agriculture, retail, human resources, and education further underscore its versatility. For instance, John Deere's use of AI in precision agriculture and Amazon's AI-powered recommendation system have set new benchmarks in their respective sectors.

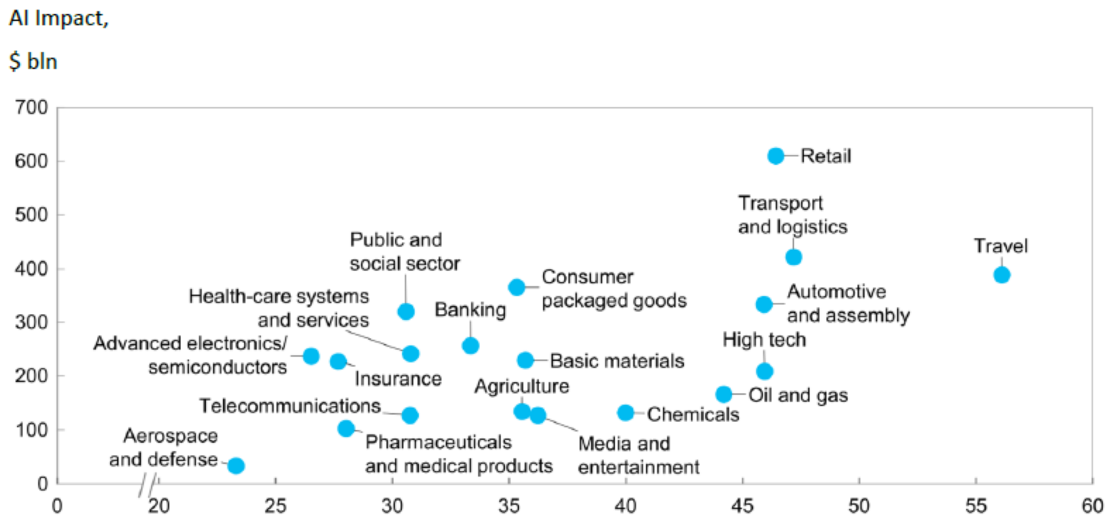
The entertainment industry, especially platforms like Netflix, personalize user experiences using AI, ensuring content relevancy. This has significantly improved user engagement and satisfaction.

According to Russell and Norvig (2010) in their seminal book, "Artificial Intelligence: A Modern Approach", the expansion of AI across industries can be attributed to its adaptability and the exponential growth in data availability and computational power. The influence of AI in sectors like energy, legal, and regulatory compliance signifies its ubiquity and indispensability in modern operations.

However, with these advancements come challenges. Gary Marcus, a prominent cognitive scientist, has often pointed out the limitations of current AI systems, emphasizing the need for more robust, generalizable solutions (Marcus, G., 2018). As industries continue to integrate AI, there's a pressing requirement for balancing innovation with ethical considerations and ensuring these systems are transparent, accountable, and beneficial to all.

McKinsey's "Notes from the AI frontier: Applications and value of deep learning" publication (2018) provided an estimation on the potential value of AI techniques. They began by defining valuable AI techniques and then explored 9 business functions across 19 industries. By identifying over 400 potential AI use cases within these sectors, they estimated the annual value each could generate, considering factors like potential revenue increase, cost reduction, and overall performance enhancement. By summing the values from all these use cases, they arrived at an estimated range of $3.5 trillion to $5.8 trillion in annual value. These figures were then benchmarked against the total potential value of all analytical techniques and were based on assumptions about factors like AI adoption rates, industry readiness, and the pace of innovation. The potential AI impact on different industries is presented in Figure 3.

Figure 3. Share in total AI impact, %



Source: McKinsey Global Institute, 2018

In summary, the evolution of AI has progressed from rule-based systems to connectionist models, machine learning, and ultimately, deep learning. These advancements have led to significant improvements in AI capabilities, resulting in applications across various industries such as healthcare, finance, transportation, manufacturing and other. Understanding the history of AI and its applications is essential for contextualizing the development and potential impact of modern AI tools like ChatGPT in the realm of software development.

1.1.2. Introduction to AI language models and their capabilities

AI language models have become an integral part of the artificial intelligence landscape, playing a critical role in understanding, interpreting, and generating human language. These models leverage natural language processing (NLP) techniques and machine learning algorithms to analyze vast amounts of linguistic data and produce human-like text based on the given input. Over the years, AI language models have witnessed significant advancements, with models such as OpenAI's GPT-3 demonstrating remarkable capabilities in generating coherent and contextually relevant text.

The primary capability of AI language models is their ability to understand and interpret context. They process large volumes of text data, identify patterns, and generate coherent and meaningful responses to various prompts or questions. According to Jurafsky and Martin (2019), "Language models give us a way to generate text that is similar to the text in a training corpus" (Jurafsky, D. et al, 2019). Some common applications of AI language models include text generation, summarization, paraphrasing, question-answering, and language translation.

Recent advancements in AI language models can be attributed to the adoption of deep learning techniques, such as the transformer architecture, which has played a crucial role in enhancing the models' understanding of context, grammar, and semantics [6]. Vaswani et al. (2017) noted that "transformer models outperform the best models on English-German and English-French translation tasks".

AI language models have demonstrated impressive performance in various tasks. As Radford et al. (2019) observed, the performance of [GPT-2] is within the range of human performance on some benchmark tests. The success of these models can be linked to their ability to learn complex patterns and structures within human language, allowing them to generate coherent and contextually appropriate text.

Moreover, AI language models can be fine-tuned for specific tasks or domains, which further enhances their adaptability and versatility. Howard and Ruder (2018) highlighted the effectiveness of transfer learning in the context of NLP, stating that "fine-tuning a pre-trained model on a downstream task can result in substantial improvements".

Table 2. Notable AI language models along with their developers and key features

Model Name	**Developer**	**Year**	**Key Features**	**Description**
Word2Vec	Google	2013	Continuous Vectors	Converts words into multidimensional vectors to identify semantic and syntactic similarities. These vectors help identify words that are similar in context (Mikolov et al, T., 2013)
ELMo	Allen Institute	2018	Contextual Word Representations	Creates word embeddings that take into account the entire context in which a word appears, improving understanding of the meanings of words in sentences (Peters, M. E., et al, 2018)
GPT	OpenAI	2018	Generative Pre-training	This model aims at generating coherent and rich text. It uses converters to improve efficiency and scalability.
BERT	Google	2018	Bidirectional Context	Revolutionizes context understanding by processing words concerning their surroundings, considering both the left and the right context in all layers, making it highly effective for various NLP tasks.
T5	Google	2020	Unified Framework	"Text-to-Text Transfer Transformer" treats every natural language processing problem as a text-to-text problem, allowing it to handle various language tasks like translation, summarization, and question-answering within a single framework (Raffel, C. et al, 2020)

Source: made by author

In conclusion, AI language models have emerged as powerful tools in the field of natural language processing, with the potential to revolutionize a wide range of applications, from content generation and customer support to research and software development.

1.2. ChatGPT: a powerful AI language model

This section highlights ChatGPT, a state-of-the-art AI language model developed by OpenAI, as a prime example of the power and potential of AI language models. We will provide an overview of ChatGPT, including its capabilities and the underlying technology that drives it. Furthermore, we will explore the many use cases of ChatGPT in different industries, demonstrating how this powerful tool can be harnessed to solve a wide array of problems, enhance productivity, and transform the way businesses operate.

1.2.1. Overview of ChatGPT and its capabilities

ChatGPT is a cutting-edge language model developed by OpenAI, based on the GPT-4 architecture. It is a powerful AI-driven tool designed to generate human-like text, understand context, and provide coherent responses to prompts or questions. ChatGPT has gained significant attention due to its impressive capabilities and potential applications across a wide array of tasks and industries (OpenAI, 2023).
It is different from the models described in the previous section by the following aspects (Henceforth Solutions, 2023):

- **Scope and functionality:**

While models like Word2Vec and ELMo are primarily focused on word embeddings and capturing semantic nuances, ChatGPT is a full-fledged text generator. It can create coherent paragraphs of text, answer questions, and even participate in conversations.

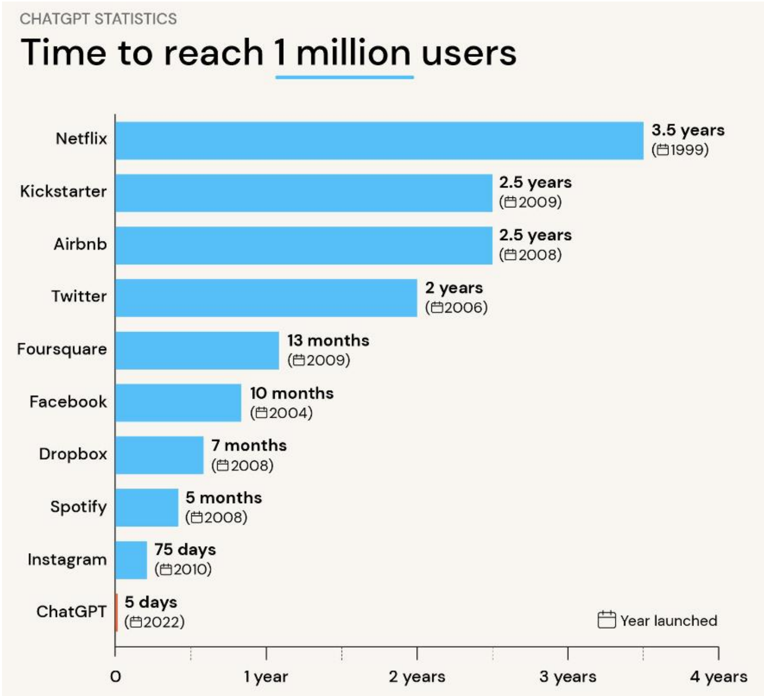
- **Generative and discriminative models:**

BERT is a bidirectional transformer used to understand context (a discriminative model), but ChatGPT is a generative model. This means that ChatGPT is good at generating text, while BERT is optimized for understanding context and is used in tasks such as text classification, object recognition, etc.

- **Unsupervised learning and flexibility:**

ChatGPT is trained using unsupervised learning on large volumes of text, allowing it to generate coherent and contextually relevant text based on patterns it recognizes from the training data.In general, the tool is able to answer questions and produce responses based on a dataset of 300 billion words and 175 billion parameters and gained 1 million users in its first week after launch.

Figure 4. An example of usage ChatGPT for content generation



Source: OpenAI

It is estimated that ChatGPT-4 trained on 100 trillion parameters, which is roughly equal to the human brain. This suggests that the training data for the latest version could be 571 times larger than the 175 billion parameters used for ChatGPT-3.

While, comparing to other types of AI it does not have a capacity now to image recognition, speech recognition, it is designed for natural language processing tasks.

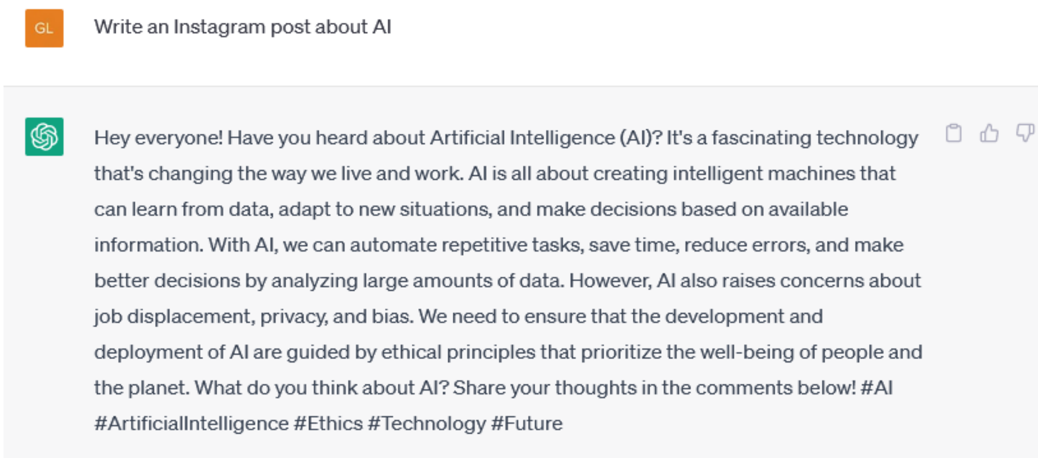
One of the primary capabilities of ChatGPT is its ability to generate contextually relevant text, making it highly suitable for tasks such as:

- Content creation: Generating articles, blogs, social media posts, and more (Picture 2);
- Summarization: Condensing lengthy documents into concise summaries;
- Paraphrasing: Rewriting text to express the same meaning in different words.

It can also be used for tasks like:

- Question-answering: Providing accurate and relevant answers to user queries;
- Language translation: Converting text from one language to another;
- Code generation: Producing functional code based on natural language inputs (Picture 3).

Figure 5. An example of usage ChatGPT for content generation



Source: made from query

Figure 6. An example of usage ChatGPT for code writing

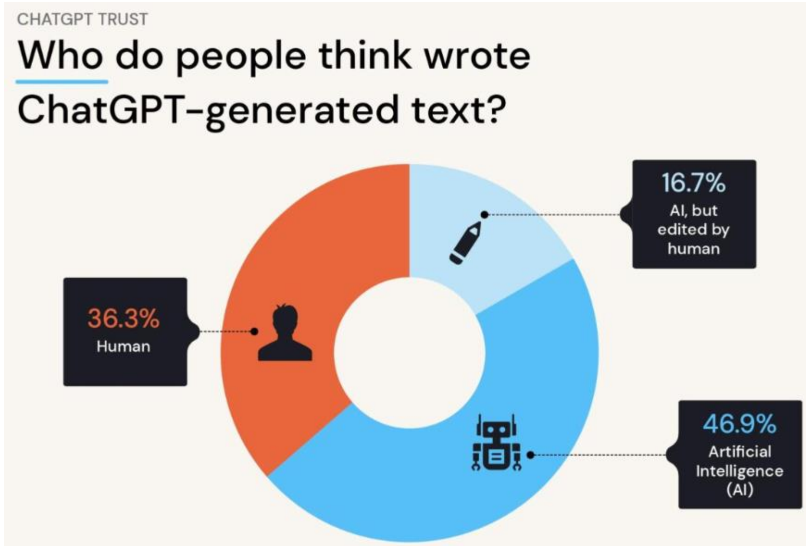
Source: made from query

Furthermore, ChatGPT can be fine-tuned to address specific tasks or domains, making it highly adaptable and versatile for various applications (Brown, T. B., et al., 2022).

The development of ChatGPT has been driven by advancements in the field of natural language processing (NLP) and deep learning, specifically the transformer architecture. As a result, ChatGPT exhibits a strong understanding of context, grammar, and semantics, allowing it to generate coherent and relevant text outputs.

Moreover, experiment wasran at Tooltester, found that more than half of readers (53%) incorrectly believed ChatGPT-generated content discussing topics such as finance, health, technology, entertainment, and travel was created by human.

Figure 7. Results of experiment



Source: Tooltester

Some notable features of ChatGPT include:

- <u>Advanced text generation</u>: Crafting text that closely resembles human writing
- <u>Context understanding</u>: Recognizing the context of a conversation or input to provide relevant responses
- <u>Large-scale learning</u>: Training on vast amounts of text data to develop a comprehensive understanding of language (Haman, M., et al, 2023).

Furthermore, AI language models have been employed in various creative applications, such as generating artwork, composing music, and producing poetry. These models' ability to generate contextually relevant content has sparked interest in their potential for augmenting human creativity across various domains.

Surely, ChatGPT is not completely perfect in giving answers yet and has its own limitations:

<u>'Hallucination'</u>: ChatGPT may produce plausible but nonsensical answers, and fixing this issue is challenging due to limitations of Reinforcement learning training (a type of machine learning where an agent learns how to behave in an environment by performing actions and receiving rewards or penalties in return). It can be difficult to design the right reward function or to gather enough appropriate feedback to guide the model effectively.

<u>Input Phrasing Impact</u>: The model's responses can vary based on input phrasing, leading to inconsistent answers.

<u>Ambiguity Handling</u>: Instead of seeking clarification for ambiguous queries, ChatGPT tends to guess user intent.

<u>Overused Phrases</u>: The model often repeats certain phrases due to training data bias.

<u>Biased Behavior and Harmful Inputs</u>: ChatGPT may exhibit bias or respond to harmful inputs, requiring moderation to mitigate risks.

To sum up, ChatGPT represents a significant milestone in AI development, offering a broad range of capabilities that can transform industries and streamline various tasks. It has its limitations, but its capabilities are broad.

1.2.2. Potential use cases of ChatGPT in various industries

ChatGPT has the potential to be applied across a wide range of industries, offering innovative solutions and support.

Table 3. Potential Use Cases of ChatGPT in Various Industries

Industry	Use Case
Content Creation	Generate articles, blogs, social media posts
Customer Support	Automate responses, assist with FAQs
Marketing	Create ad copy, analyze audience sentiment
Education	Tutoring, generate study materials
Legal	Draft contracts, analyze legal documents
Healthcare	Diagnose illnesses, offer personalized care advice
Translation	Convert text between languages
Software Development	Generate code, debug programs
Finance and Investment	Analyze financial data, generate investment insights
Human Resources	Screen resumes, draft job descriptions
Gaming and Entertainment	Create engaging narratives, personalize storylines

Source: made by author

The table above provides a concise summary of the potential use cases for ChatGPT across various industries.
It is evident that ChatGPT's capabilities can be harnessed to solve diverse problems and enhance productivity in a multitude of sectors. AI language models like ChatGPT are revolutionizing the way businesses and organizations operate by offering innovative solutions to complex problems. These models allow for the automation of various tasks, enabling professionals to focus on more strategic aspects of their work. Moreover, AI language models can adapt to specific domains, making them highly versatile tools for different industries.
As research in this area continues to grow, it is likely that the range of applications for ChatGPTs will expand, offering even more opportunities for innovation and problem-solving.

1.3. Productivity in Software Development

In this section, we examine the concept of productivity within the context of software development, a critical aspect for businesses and organizations striving for efficiency and competitiveness in today's technology-driven world. We will discuss the importance of productivity in the software development process and explore various metrics used to measure it. Then we will smoothly go to analyzing the methodology and results of existing researches on AI influence on productivity and will analyze on of the first studies on affects on productivity of ChatGPT. Through this chapter, we aim to provide a comprehensive understanding of AI's influence on. Understanding productivity is essential for identifying areas of improvement and harnessing the potential of innovative technologies, such as AI and ChatGPT, to optimize the development process.

1.3.1. Defining productivity and its significance

Productivity is a fundamental concept in economics and business. It is often defined as "the ratio between the output volume and the volume of inputs" (Paul Krugman, 1994).
Solow (1957) further elaborates on productivity, stating that technical change and the aggregate production function are key drivers of productivity.
The significance of productivity is multifaceted, as it directly impacts economic growth, competitiveness, and living standards. Syverson (2011) says "productivity is efficiency in production: how much output is obtained from a given set of inputs. As such, it is typically expressed as an output–input ratio" and emphasizes the importance of productivity, arguing that variations in productivity levels are crucial in understanding differences in living standards and growth rates across countries and over time.
In addition, productivity plays a pivotal role in determining the competitive advantage of businesses. As Porter (1990) asserts, firms that achieve high levels of productivity enjoy a competitive advantage in the marketplace, as they can offer better products or services at lower prices.
Productivity is a key factor in determining project success, as it affects cost, schedule, and quality. Also, increasing productivity in software development can lead to reduced costs, increased profits, and improved customer satisfaction.
Paul Krugman, a renowned economist, also highlights the significance of productivity. In one of his famous quotes, Krugman (1994) stated "productivity isn't everything, but in the long run it is almost everything. A country's ability to improve its standard of living over time depends almost entirely on its ability to raise its output per worker".
In summary, productivity is a vital concept in economics and business, with direct implications for economic growth, competitiveness, and living standards. Understanding and enhancing productivity is crucial for businesses and economies alike to maintain a competitive edge and improve overall well-being.

1.3.2. Metrics for measuring productivity

To effectively measure productivity, various metrics have been proposed and utilized across different industries. Syverson (2011) emphasizes that some of the widely recognized productivity metrics include:

- Output per hour worked: This metric, often used in the context of labor productivity, measures the amount of goods and services produced for each hour of labor. Baily and Gordon (1988) state that output per hour worked has been the most common measure of productivity growth in economic studies.
- Total Factor Productivity (TFP): TFP measures the efficiency with which inputs (e.g., labor, capital) are transformed into outputs. Hulten (2001) notes that it is often

considered a better indicator of long-term economic growth than output per hour worked, as it accounts for changes in technology, management practices, and other factors affecting productivity. According to Solow (1957), "TFP growth is the portion of output not explained by the amount of inputs used in production".

- Value Added per Worker: It calculates the value-added contribution of each employee to the production process. Nishimizu and Page (1982) explain that "value added per employee is a useful measure of productivity because it reflects the contribution of labor to the production of goods and services, net of the cost of intermediate inputs".
- Return on Investment (ROI): ROI is a financial metric that evaluates the efficiency of an investment, comparing the net profit to the initial investment. ROI can be used to measure productivity in various industries, as it provides insights into how effectively resources are being utilized to generate returns.

These metrics, among others, can provide valuable insights into productivity across a wide range of industries and help organizations identify areas for improvement.

1.3.3. Review of survey-based studies on AI and productivity

Having reviewed the primary techniques for assessing productivity, our focus now shifts towards analyzing articles that have evaluated the influence of artificial intelligence on productivity.
We found only a few articles and carefully studied three of them:

- Damioli et al, 2021, The impact of artificial intelligence on labor productivity;
- Czarnitzki et al, 2022, Artificial Intelligence and Firm-Level Productivity;
- Alderucci etl al, 2019. Quantifying the Impact of AI on Productivity and Labor Demand: Evidence from U.S. Census Microdata.

1. Damioli et al, 2021, The impact of artificial intelligence on labor productivity

This article to study productivity uses number of patents as a method. They use a database of 5257 AI patenting firms to evaluate the short-term effect of AI technologies on firm labor productivity took. Authors control for firms' patenting activities in AI and non-AI related fields together with accounting information (including turnover, employment, and capital formation), country location, and industrial activity and use a worldwide sample of companies from four continents that have filed at least one patent related to the field of AI between 2000 and 2016.
They define labor productivity as turnover/number of employees. The key explanatory variable of interest is AI patent applications, which measures the change in a firm's knowledge stock in the field of AI. In addition to the number of AI patent applications as proxy for a firms' knowledge stock, they also took into account innovative efforts in non-AI related fields. Other variables included in the dynamic productivity model are the growth in employment, measured as the number of employees expressed in full-time equivalents, and the growth in the capital stock, approximated by the growth in fixed capital. The models control for firm size (employment), industry, year and country-specific differences in labor productivity dynamics.

The main conclusion is that, once controlling for other patenting activities, AI patent applications generate an extra-positive effect on companies' labor productivity. Smaller, more agile AI-patenting firms may have been able to readjust faster and introduce AI-based applications in their production processes at a scale allowing the creation of a significant impact on productivity. (Damioli et al, 2021).

2. **Czarnitzki et al, 2022, Artificial Intelligence and Firm-Level Productivity**

The authors point out that there are only very few studies investigated likely productivity effects of AI at the firm-level, mentioning the reason, "presumably because of lacking data". They exploit unique survey data on firms' adoption of AI technology and estimated its productivity effects with a sample from the German part of the European Commission's Community Innovation Survey (CIS). The information collected is representative for all firms in Germany with at least 5 employees in manufacturing, mining, utilities, and business-oriented service sectors (wholesale trade, transportation, financing and insurance, information and communication, professional, scientific, technical, administrative and support services).

As a method, cross-sectional dataset and a panel database were used. To overcome potential endogeneity issues of AI use, they employed instrumental variable regressions using AI diffusion at industry level, the firm's past investment in R&D and innovation, and organizational rigidities as instruments.

They followed the standard approach to analyze firm productivity by linking inputs and outputs within a production function (The production function (f) of firms describes the association between a firm's output (Y), measured by annual sales, and total factor productivity (A) as well as a set of inputs, such as capital (K), labor (L), and intermediate inputs such as materials, energy and purchased services (M). However, they accommodated this framework and added an additional input to the production function that represents AI adoption (AI).

As a result, the authors found positive and significant effects of the use of AI on firm productivity. This finding holds for different measures of AI usage, i.e., an indicator variable of AI adoption, and the intensity with which firms use AI methods in their business processes.

3. **Alderucci etl al, 2019. Quantifying the Impact of AI on Productivity and Labor Demand: Evidence from U.S. Census Microdata**

The authors use U.S. Census Microdata and machine learning algorithms to assess the degree to which patent grants are AI-related.

The study uses U.S. Census Microdata collected on the innovating firms as a sample. The authors match their data on AI patenting to this sample and perform an event study using these matched data to gauge the impact of AI-related innovations on firm labor demand and firm growth. The sample includes a wide range of firms across different industries that have been granted patents related to AI.

The study uses a standard Cobb-Douglas production function to measure the effects of AI-related innovations on productivity and labor demand. The authors introduce counts of AI patents or a dummy variable equal to 1 when AI patenting begins as a separate regressor. They also use an event study analysis to better control for endogeneity and look at firm behavior and outcomes before and after the AI innovation. The variables used in the event study analysis include within-firm changes to revenue and value-added resulting from

innovations in AI, as well as firm behavior and outcomes before and after the AI innovation. The identification relies on matching each firm with at least one AI-related patent as closely as possible with a similar same-industry counterpart which does not obtain an AI-related patent.
The authors highlight that AI-related innovations appear to raise output per worker and increase within-firm wage inequality. They claim that the impact of AI on productivity is concentrated in a small number of leading firms rather than being broadly observed across all firms. The study finds that AI-related patents have a positive impact on productivity, but this impact is concentrated in a small number of leading firms rather than being broadly observed across all firms.

Table 4. Summary of general information on the following articles

Article	Methodology	Sample	Key Findings
Damioli et al, 2021, The impact of artificial intelligence on labor productivity	Uses number of patents as a method; controls for various firm, industry, and country factors	5257 AI patenting firms from 2000 to 2016	AI patent applications have a positive effect on labor productivity, especially in smaller, agile firms
Czarnitzki et al, 2022, Artificial Intelligence and Firm-Level Productivity	Uses cross-sectional and panel databases; instrumental variable regressions	German firms from the European Commission's Community Innovation Survey	Positive and significant effects of AI use on firm productivity; results hold for various measures of AI usage
Alderucci et al, 2019, Quantifying the Impact of AI on Productivity and Labor Demand	Uses U.S. Census Microdata and machine learning algorithms; event study analysis	U.S. firms with AI-related patents	AI-related innovations positively impact firm growth, productivity, and within-firm wage inequality

Source: made by author

In conclusion, the review of survey-based studies on AI and productivity reveals a variety of methodological approaches used to investigate the impact of AI on labor and firm-level productivity. Despite the differences in data sources, samples, and research designs, a consistent theme emerges from these studies: AI technology has a significant positive impact on productivity across industries and regions.
Firms that adopt AI technologies consistently experience increased growth in employment, revenue, and output per worker, regardless of the specific method used to measure productivity. Although some studies point out that the impact of AI on productivity might be more pronounced in a small number of leading firms, the overall consensus across these diverse approaches highlights the transformative role AI can play in enhancing productivity.

1.3.4. Review of empirical analysis of ChatGPT's impact on productivity

As of our literature review that was made in May 2023, we came across only one scientific article containing empirical data that thoroughly examined the effects of ChatGPT on productivity. Given its potential to offer insights into research methodology and serve as a foundation for result comparison, we deemed it crucial to conduct a detailed analysis of this article.

The article is made by Noy et al and published in March 2023 under the name Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. It represents the online experiment, the researchers recruited 444 experienced, college-educated professionals from various occupations such as marketers, grant writers, consultants, data analysts, human resource professionals, and managers. Each participant was assigned two occupation-specific, incentivized writing tasks, including press releases, short reports, analysis plans, and delicate emails. These tasks were designed to resemble real tasks performed in their respective occupations and took 20-30 minutes to complete.

Participants were randomly divided into two groups:

the first group was called a treatment group (approximately 50% of participants). They were said to use ChatGPT between their first and second tasks;

the second group is the control group. They were said to use LaTeX editor Overleaf.

Through the use of both within-person and between-person variance, the researchers were able to quantify the causal effects of ChatGPT using this approach. Analysis of inequality was made possible by using performance on the first task as a baseline indicator of aptitude.

The participants' output, overall time spent on the work, time spent on its various subcomponents, job satisfaction, self-efficacy, and beliefs regarding automation were all collected by the researchers. In order to create an objective measure of time spent on the task and to identify ChatGPT usage in the control group and on the pre-treatment task, they also recorded pictures of each participant's output every minute while they were working on the task. Blinded experienced individuals from the same professions who were motivated to assign high grades to outputs evaluated the outputs' quality.

These are the key findings:

Take-up of ChatGPT: 92% of treatment group participants signed up, with 81% using it on the second task.

Productivity (earnings per minute): The treatment group saw a 37% decrease in time taken and a 0.45 standard deviation increase in evaluator grades.

Supplementary interventions: One supplementary intervention, which required participants to spend exactly 15 minutes on each task, showed a similar increase in grades by 0.39 standard deviations in the treatment group. Another intervention allowed participants to edit their first-task output using ChatGPT, with 23% choosing to replace their response and 25% using ChatGPT to edit their original response. This suggests that participants view ChatGPT as a way to improve output quality and save time.

Productivity Inequality: In the control group, participants' average grades on the first and second tasks had a correlation of 0.49, indicating that there was ongoing productivity inequality. Initial disparities were partially eliminated in the therapy group, with a 0.25 correlation between first-task and second-task grades. Due to the fact that individuals with lower first-round scores benefited more from ChatGPT access, there was a decrease in inequality.

Task Structure: ChatGPT changed the structure of writing tasks, reducing time spent on drafts and increasing time spent on editing.

Skill Demand: Tests revealed no clear evidence that ChatGPT is particularly helpful for those with poor writing skills compared to others. Both willingness to pay for ChatGPT and grade gains from its use were roughly flat across different levels of writing skills.
Self-Efficacy and Job happiness: ChatGPT slightly and imprecisely raises self-efficacy by 0.20 standard deviations and raises job happiness by roughly 0.40 standard deviations. Many participants took pleasure in learning about and using the instrument.
Beliefs About Automation: Participants' beliefs about automation grew by 0.20 standard deviations in net optimism, 0.39 standard deviations in excitement, and 0.26 standard deviations in fear after using ChatGPT.
Two-Week Follow-Up Survey: Compared to the control group participants (18%), 33% of former treatment group participants reported using ChatGPT at work in the previous week. Its average usefulness rating from users was 3.65 out of 5. Most people who do not use ChatGPT at work complained that it lacked the contextual information necessary for their writing assignments. However, the result that ChatGPT can boost productivity on many mid-level professional writing tasks is supported by the fact that many respondents choose to use it in their actual work.
To sum up, this study examined the impact of ChatGPT on various aspects of professional writing tasks, including productivity, task structure, skill demand, job satisfaction, self-efficacy, beliefs about automation, and productivity inequality.
The results demonstrated that ChatGPT significantly improved productivity by reducing time taken and increasing evaluator grades. It also changed the task structure by reducing time spent on drafts and increasing editing time. Although no clear evidence was found that ChatGPT is particularly helpful for those with poor writing skills, it did increase job satisfaction and self-efficacy, as well as influenced beliefs about automation.
Furthermore, ChatGPT reduced productivity inequality by benefiting participants who scored lower on the first round. The two-week follow-up survey revealed that a significant portion of participants continued to use ChatGPT in their jobs, supporting its potential to increase productivity in various mid-level professional writing tasks.
The following results are thought-provoking and will be used while designing our own study.

1.3.5. Measuring productivity in Software Development

The significance of productivity in software development cannot be overstated, as it directly impacts the organization's ability to deliver software projects on time, within budget, and of high quality.
According to Banker and Kauffman (1991), software productivity can be found from the following formula:
Productivity = (Size of Application Developed) / (Labor consumed during development).
According to Wagner and Ruhe (2008), software productivity can be measured traditionally using the lines of code or function points and the productivity is the LOC or FP produced per hour by the programmer.
Sudhakar, G. P. et al (2011) made an overview of techniques/models for measuring software development productivity (Table 5).

Table 5. Techniques / Models for measuring Software Development Productivity

Sl. No:	Technique/ Model	Formula/Description	Highlights	Reference
1.	Team Productivity (P)	P = Kilo Lines of Code/ Person months of effort	Further given the needed staff size as "person months of effort divided by project time duration in months"	Tausworthe (1982)
2.	Measurement model	Analysis/Design Activity Output measure = Function Points Coding/Testing Activity Output measure = Source Lines of Code Input Measure = Total Labor hours	This model considers Function Points, SLOC, environmental variables, and any deviations from the project.	Banker, Datar and Kemerer (1991)
3.	Productivity Model and Cost Model	Mathematical Models	This Model explains the impact of interaction of team members and team size on team productivity and project cost.	Tockey (1996)
4.	Model of Life Cycle Productivity and Quality	Quality = f1(Personnel Capability, Usage of Tools, Product Size in LOC, PROCESS, Front End Resources) Life Cycle Productivity = f2(Conformance Quality, Personnel Capability, Usage of Tools, PROCESS) Life Cycle Productivity = Product size in LOC / Total cost incurred in Product development and support.	This model considers variables such as personnel capability, quality, software process, product size in LOC, Front End Resources and Usage of tools.	Krishnan, Kriebel, Kekre and Mukhopadhyay (1999)
5.	Model of Correlated Team Behavior	Software Team productivity = KLOC per Calendar month.	Provides a simulation model which supports correlated team behavior.	Potok and Vouk (1999)

6.	Productive Ratio (J)	J = % of Direct Development time / % of Idle time	The model suggested considered productivity, requirements volatility and complexity.	Nogueira, Luqi, Berzins and Nada (2000)
7.	Productivity Model	Productivity = Number of Function Points / Effort in Man months	This model considers the factors such as Experience of Project Manager, size, requirements ambiguity, complexity, stable standards, user requirements, usage of tools, etc.	Blackburn, Lapre and Van Wassenhove (2002)

Source: Sudhakar, G. P. et al., 2011, Serbian Journal of Management

These techniques/models use different formulas and descriptions to measure productivity based on various factors such as lines of code (LOC), person months of effort (PMOE), function points (FP), object points (OP), use case points (UCP), feature points (FP), and quality metrics. The highlights for each technique/model provide a brief summary of its strengths or unique features. For example, the Measurement Model considers environmental variables and deviations from the project while calculating productivity measures.

In general, productivity in software development can be measured quantitatively:

- Lines of Code (LOC): This metric measures the size of the software in terms of the number of lines of code written. It is a simple and widely-used metric, but it has limitations, as it does not account for the complexity of the code or the varying productivity levels of different programming languages.
- Function Points (FP): they evaluate the quantity of inputs, outputs, enquiries, internal and external interfaces, and files to determine the software's functionality. Because it takes into account the value provided to the end-user rather than merely the quantity of the code, this statistic offers a more realistic depiction of productivity.
- Object Points (OP) are a unit of measurement for the size and complexity of object-oriented software. This statistic accounts for the quantity, complexity, and interrelationships of the classes. The productivity of projects that make use of object-oriented programming languages can be assessed using this method very well.

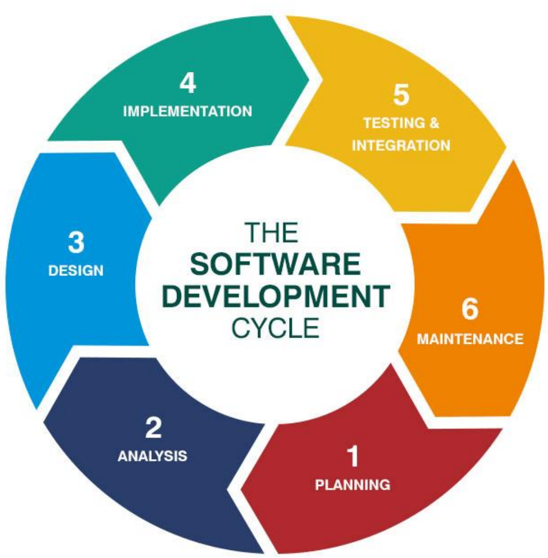
However, Tomaszewski' P. (2006) in the dissertation "Software Development Productivity Evaluation and Improvement for Large Industrial Projects" points out that "despite a relatively simple equation (product size/development effort) and an easy-to-grasp meanings, the application of the productivity metric to software development is not straightforward and standardized. We must be very careful when comparing productivity between different projects. The important thing is to assure that we compare the same things.

For example, in one project the effort metric may include only the hours spent by the designers and testers, while in the other one it may contain the work hours of designers, testers, managers, and technicians. Comparing the productivity of these two projects using their understanding of the effort will not give any meaningful results".
Since in our study we will be assessing the impact on productivity in different companies and teams, we will not continue further consideration of approaches to quantitative assessment, since we will be comparing heterogeneous things, respectively, these approaches cannot be applied.

1.4. Software Development Process and ChatGPT's transformational influence

In this section, we will explore the impact of AI on productivity in software development, with a special focus on ChatGPT. We will begin by discussing how AI is transforming the software development process, touching upon the various stages of the process and examining the role AI plays in each. We will continue, understanding first results of using ChatGPT.

Before discussing how ChatGPT is transforming the software development process, it is essential to understand what "software development process" means and its different stages. Institute of Electrical and Electronics Engineers (the world's largest technical professional organization) defines software engineering as "the application of a systematic, disciplined, quantifiable approach to the development, operation and maintenance of software.
Klopper R. et al (2007) outlines that all software projects go through the following stages of the software development process (Picture 4):
Figure 8. SDLC



Source: Pressman, R. S. (2010). Software engineering: A practitioner's approach (7th ed.). McGraw-Hill.

The essential stages of the software development process include:

1. Planning: this stage involves defining the project's scope, objectives, and constraints. It includes identifying stakeholders, establishing communication channels, and determining resource requirements. Project managers and stakeholders collaborate to create a project plan that outlines tasks, timelines, and budgets. ChatGPT can be utilized in planning and project effort estimation by assigning staff to tasks based on their experience and ability, determining the relationships between tasks, and estimating task durations in a way that meets the project completion date (Padmanaban et al, 2019). According to a study by Davenport and Ronanki (2018), ChatGPT can improve project planning through intelligent forecasting, which leads to more accurate estimations of project completion times and budgets).
2. Analysis: During the analysis stage, the project team gathers and analyzes requirements from stakeholders to understand the problem they're trying to solve. It can assist in requirements analysis by identifying patterns in data, extracting insights, and suggesting improvements based on historical data. According to Nair et al. (2018), ChatGPT can contribute to automating the requirements elicitation process, leading to better understanding and communication between stakeholders (Kumari et al, 2018).
3. Design: the phase involves creating a blueprint for the software solution based on the requirements gathered during the analysis stage. Software architects and designers collaborate to create high-level and detailed designs, including system architecture, user interfaces, and data models. Leveraging ChatGPT's capabilities, designs can be generated and optimized by analyzing existing system models, ensuring robustness and quality.
4. Implementation: this is where the actual coding takes place. ChatGPT enhances this step by suggesting code completions, predicting snippets, and quickly identifying bugs or vulnerabilities (Kumari et al, 2018).
5. Testing and integration: this stage ensures functionality, performance and compatibility of the software. ChatGPT can help to automate tests or to write a test cases strategy (Raroque Ch.).
6. Maintenance: during post-release, ChatGPT serves as a technical support tool, providing answers to developer queries, offering troubleshooting assistance, and recommending best practices. Furthermore, its documentation capabilities can ensure that software enhancements are well-documented for future reference.

Chapter 2. Methodology and Approach

This section provides a description of the main goal and design of the study, a description of the main stages of the research process with a detailed explanation of the development of a mixed methodology, within which data for analysis will be collected, and this data will be processed by the selected tools in order to answer the research question, tested all scientific hypotheses and the goal of the study has been achieved. In addition, limitations of the data and methods used are presented at the end of the chapter. The primary research aim is to provide valuable insights of how ChatGPT influenced productivity of software development on different stages. Overall, this section provides a compelling rationale for the selection of relevant data and methods for the study.

2.1. Research hypotheses

The rapid growth of artificial intelligence technologies, particularly AI-driven language models like ChatGPT, has garnered considerable interest in their potential applications in software development. However, the impact of ChatGPT on productivity in software development remains underexplored. Identifying the extent of ChatGPT's influence on productivity can provide valuable insights for organizations seeking to optimize their processes and drive innovation in the competitive landscape.

As was mentioned before, the main goal of this study is to investigate the influence of ChatGPT on productivity in the software development process and provide valuable insights on its impact.

The study will test the following hypotheses and will expect the following results (Table 6):

Table 6. Hypothesis to test and expected results

Hypothesis	Expected Result	Source
H1.Less-experienced employees benefit more from using ChatGPT than those with higher levels of expertise.	Higher ChatGPT impact scores among respondents with less experience in the software industry.	Junior developers can engage with ChatGPT for guidance, best practices, or to understand complex algorithms, ensuring that they ramp up quickly (Kumari et al, 2018)
H2. Employees having a lower grade benefit from ChatGPT more.	Higher ChatGPT impact scores among respondents with fewer grade (Junior) comparing to Middle and Senior.	Junior developers can engage with ChatGPT for guidance, best practices, or to understand complex algorithms, ensuring that they ramp up quickly (Kumari et al, 2018)

H3. There will not be a difference in influence on Productivity for employees work in large organizations or for employees of small firms / freelancers.	Same ChatGPT impact scores among respondents from different categories of company size.	Interviews with developers
H4. ChatGPT's impact will vary for different software development stages.	Different ChatGPT impact scores for different stages.	The section "Understanding Software Development process Software Development Process and ChatGPT's transformational influence" and (Pressman, R. S. (2010)
H5. The impact of ChatGPT on software development productivity is more significant in less creative stages.	Higher ChatGPT impact scores in less creative stages compared to more creative stages.	ChatGPT enhances this step by suggesting code completions, predicting snippets, and quickly identifying bugs or vulnerabilities (Kumari et al, 2018)
H6. Professional using paid version of ChatGPT (4.0) will record higher positive influence on their productivity.	Higher overall impact on productivity score given by participants that have a paid version of ChatGPT.	Interviews with developers
H7. Younger participants will record more impact of ChatGPT as they tend to implement new technologies to life more actively.	Higher overall impact on productivity score given by younger participants.	Interviews with developers
H8. The impact of ChatGPT on software development productivity is higher in teams where the usage of ChatGPT is actively promoted by line manager (however, we supposed that the active promotion is rare).	Higher ChatGPT impact scores among respondents whose line manager is aware and supportive of ChatGPT usage.	Interviews with developers

Source: made by author

To create the hypotheses for our study, we began by brainstorming ideas based on reviewed literature and our knowledge of ChatGPT and its potential applications in the software

development industry. We considered various factors that could influence the effectiveness and utility of ChatGPT, such as experience levels, job roles, and specific stages of the software development process.
In addition to our initial ideas, we conducted informal interviews with 3 software developers who had experience using ChatGPT in their work. We asked them about their impressions of the tool, how they use it in their daily tasks, and whether they observed any specific benefits or drawbacks. Their insights provided valuable input that informed and refined our hypotheses.
By combining our initial ideas, literature review and the insights from the software developers we spoke to, we were able to develop a set of hypotheses that capture various aspects of ChatGPT's potential impact on the software development industry. These hypotheses will guide our research and provide a framework for analyzing the data collected through our survey.

2.2. Research design

The research design includes a survey targeting different professionals involved in the software development process. The survey consists of various question types, designed to test each hypothesis (Table 7).
Table 7. Mapping of Question Types and Hypothesis Tested

№	Question Type	Hypothesis Tested
1	Demographic questions (age, gender, years of experience, job role, position level, employment type, company size, country of operations, and company's main industry)	H1, H3, H7
2	Closed-ended questions (use of ChatGPT, frequency of use)	H5
3	Likert scale questions (perceived impact on productivity, task completion speed, work quality, error reduction, problem-solving abilities)	H1, H2, H4, H5, H6
4	Stage-specific Likert scale questions (impact of ChatGPT on each stage of software development)	H4, H6
5	Yes/No question (line manager's awareness of ChatGPT usage)	H8
6	Multiple-choice question (most 'creative' stages)	H5

Source: made by author
The survey comprises a variety of question types to ensure comprehensive data collection for hypothesis testing. The questions and their respective types are described below:

1. Demographic Questions

These questions gather information on the participant's background, which can be used to identify trends and patterns among different subgroups.

- Age (open-ended);
- Gender (multiple choice: Male, Female, Other or prefer not to say);
- Years of experience in the software development industry (open-ended);
- Current job role (multiple choice);
- Current position level (multiple choice: Intern, Junior, Middle, Senior);
- Employment type (multiple choice: Regular, Freelancer);
- Company size (open-ended);

- Country of operations (multiple choice);
- Company's main industry (multiple choice).

2. Closed-ended Questions

These questions collect information on specific aspects of the participant's experience with ChatGPT.

- How frequently do you use ChatGPT for work? (multiple choice: Every day, 3-4 days a week, 2 or less days a week, several times a month, stopped using it at all). Here we should say that we did not ask to participate in the survey professional who never tried to use ChatGPT.

3. Likert Scale Questions

These questions measure the participant's perception of ChatGPT's impact on various aspects of their work.

- Overall impact of ChatGPT on productivity (1 = No impact, 5 = Extremely impactful);
- ChatGPT has helped me to complete tasks more quickly (1 = Strongly disagree, 5 = Strongly agree);
- ChatGPT has improved the quality of my work (1 = Strongly disagree, 5 = Strongly agree);
- ChatGPT has reduced the number of errors in my work (1 = Strongly disagree, 5 = Strongly agree);
- ChatGPT has enhanced my problem-solving abilities (1 = Strongly disagree, 5 = Strongly agree).

4. Stage-specific Likert Scale Questions

These questions evaluate the perceived impact of ChatGPT on different stages of the software development process.

- Impact of ChatGPT on Planning, Requirements analysis, Software design, Coding, Code review, Testing and QA, Deployment, Maintenance and bug fixing, and Documentation (1 = No impact, 5 = Extremely impactful).

5. Yes/No Question

This question investigates the awareness of the participant's line manager regarding their usage of ChatGPT.

- Does your line manager know that you are using ChatGPT for work? (Yes/No/Don't want to tell).

6. Multiple-choice question (most 'creative' stages)

This question aims to gather information on which stages of the software development process the participant considers the most creative.

- Name 4 or fewer stages that you consider the most 'creative' (multiple-choice).

These question types, including demographic, closed-ended, Likert scale, and multiple-choice questions, ensure a comprehensive and balanced approach to collecting data from the participants, allowing for effective hypothesis testing and analysis of trends and patterns within the software development industry.

2.3. Sampling Strategy and Data Collection

We adopted a purposive sampling strategy for this study to target software development professionals who had experience using ChatGPT. The primary objective of the sampling strategy was to ensure a diverse representation of perspectives from participants belonging

to different industries, levels of experience and involved in different stages of software development.
To achieve this, we used multiple platforms for data collection, including LinkedIn, Telegram, and WhatsApp.
Specifically, in Telegram and WhatsApp, we leveraged professional groups focused on software development, such as "IT Relocation Germany," "Software Development club of Higher School of Economics Alumni", etc.
This allowed us to engage with individuals actively involved in the software development community and who had insights to share regarding ChatGPT's impact on productivity.
On LinkedIn, we identified potential participants by searching for individuals who listed "software development" as their industry. By targeting individuals with diverse job roles and experiences, we aimed to gather a comprehensive range of perspectives on ChatGPT's influence on productivity in different stages.
The data collection period began in March and ended at the end of July. We sent more than 3000 messages to prospective participants inviting them to take part in the survey.
As a result of our efforts, we received 150 responses from professionals willing to share their experiences and insights on the topic.

2.4. Data Analysis

2.4.1. Method's description

The collected survey data were carefully analyzed with such statistical techniques such as correlation analysis, analysis of variance (ANOVA), means description.
Correlation analysis was used to explore the relationships between various continuous variables, ex: the duration of ChatGPT usage, its impact on productivity, and other relevant factors. By calculating correlation coefficients, we assessed the strength and direction of these relationships, providing valuable insights into how different variables are associated.
Analysis of variance (ANOVA) was employed to investigate the impact of ChatGPT across multiple groups, such as different job roles or industries. By analyzing variance between groups, we can identify if there are significant differences in productivity outcomes based on these categorical factors.
Additionally, we used means, modes, medians to understand the difference of impact on different software development stages.

2.4.2. Justification

For this investigation, a quantitative method was adopted for a number of reasons. First off, researchers can gather big enough sample sizes compared to qualitative methods. Additionally, the results are thought to be less skewed because it is relatively simple to compare them to other similar research, should they be released in the future. Additionally, in light of the research tools and study goals, this strategy is the best option. The fundamental data in this white paper are gathered and further examined.
Survey is the main method used in current research to get the relevant data and information.

When creating this format, attention was paid to simplicity and clarity, and the Likert scale was adopted as a main method, as we said earlier, Tomaszewski' P. (2006) in the dissertation "Software Development Productivity Evaluation and Improvement for Large Industrial Projects" points out that "we must be very careful when comparing productivity between different projects. The important thing is to assure that we compare the same things". We decided that since we are taking into consideration different industries, companies, tasks etc., it would be better not to use such metrics as "lines of code per hour" and etc. as the scope of work is not consistent.
Likert scale, instead, allows to give relative estimations. The value of this scale is in the range of rank (5) (Strongly agree) and rank 1 (Strongly disagree), that is, 5 degrees, and this approach is widely used in management science.

2.5. Validity and Reliability

To ensure the validity and reliability of our research findings, we have taken several measures, including:

- Designing the survey with clear, concise, and relevant questions to accurately capture the opinions of respondents;
- Employing a purposive sampling strategy to target software development professionals with experience using ChatGPT;
- Ensuring the anonymity of respondents to encourage honest and unbiased responses;
- Using rigorous statistical techniques, such as ANOVA and correlation analysis, to identify meaningful patterns and relationships in the data.

2.6. Ethical Considerations

The ethical standards for research involving human subjects are followed in this work. All survey participants provided their informed consent, and their identity and confidentiality were upheld throughout the process of gathering and analyzing the data. To protect the privacy of all participants, the research findings will be reported without revealing any personally identifying information.

In summary, this research methodology aims to provide a comprehensive understanding of the impact of ChatGPT on productivity in software development by collecting and analyzing data from experienced professionals in the field.

2.7. Limitations of the Research Method and Survey

Even though our research methodology and survey design have been carefully developed to investigate the impact of ChatGPT on productivity in software development, there are some limitations that may affect the generalizability and applicability of the findings. In this section, we discuss these limitations and their potential implications for the study.

2.7.1. Sample Size and Representativeness

The sample size of this study may be not large enough to provide a fully representative picture. Additionally, the purposive sampling strategy may have selection bias, as participants with particular characteristics or experiences might be more likely to respond to the survey. This all can potentially limit the generalizability of the findings to the broader population of software development professionals.

2.7.2. Self-reported Data

The survey relies on self-reported data from respondents, what can introduce biases and inaccuracies. Participants may be influenced by a bias of social desirability, that may lead them to provide responses they believe are expected or acceptable rather than their true opinions. Additionally, respondents may have different interpretations of the survey questions or may not accurately recall specific details related to their experiences with ChatGPT.

2.7.3. Cross-sectional Design

The cross-sectional design of the survey captures a snapshot of participants' experiences with ChatGPT at a specific point in time. As a result, the findings may not reflect changes in the impact of ChatGPT on productivity over time, or the dynamic nature of software development processes and technologies. Longitudinal research would be required to track the evolution of ChatGPT's influence on productivity and understand the long-term implications of its implementation.

2.7.4. Potential Confounding Factors

There may be confounding factors that influence the relationship between ChatGPT usage and productivity in software development, which are not controlled for in the survey. For example, differences in organizational culture, management practices, or access to resources could affect the extent to which ChatGPT is integrated into the software development process and its resulting impact on productivity. Without controlling for these factors, it is challenging to establish a direct causal link between ChatGPT usage and productivity.

2.7.5. Subjectivity in Data Analysis

The subjective opinions and presumptions of the researchers may have an impact on how they interpret survey data and identify patterns and linkages. Even though statistical tools like t-tests and correlation analysis might lessen subjectivity in data analysis, there is still a chance that the results could be skewed by the researchers' preconceived notions or expectations.
In conclusion, it is essential to recognize and acknowledge the limitations of the research method and survey employed in this study. These limitations should be considered when interpreting the findings and drawing conclusions about the impact of ChatGPT on productivity in software development. Future research may address these limitations by employing alternative methods, such as longitudinal studies or experimental designs.

Chapter 3. Findings

In this section, we will go through the most exciting part – analysis of the results. We will assess in what extent and in which ways how ChatGPT influences productivity in different stages of software development, investigate correlations between ChatGPT's influence and respondent characteristics, see the results of ANOVA tests to identify differences in ChatGPT's influence on productivity among different respondent groups and examine which stages benefit most from ChatGPT during the software development process.

3.1. Description statistics

In order to start working with our dataset, we cleaned it and gave variable names to our questions. In Appendix 2 you may find the result of this procedure. Let's have a look at the data we got.

3.1.1. Age group

Table 8. Age group

	Counts	Total	Proportion
18-24	36	150	0.240
25-34	84	150	0.560
35-44	28	150	0.187
45 and above	2	150	0.013

Source: dataset analysed by author in JASP

The age range of the participants is 18 to 54, with a 29.5 year average. The bulk (56%) of the population is in the 25–34 age range, with 18–24 (24%) and 35–44 (19%) following. Two persons (1.3%) are over the age of 45.

3.1.2. Gender

Table 9. Gender

	Counts	Total	Proportion
Female	43	150	0.287
Male	107	150	0.713

Source: dataset analysed by author in JASP

Most respondents identify as male (71.3%), while a smaller proportion identifies as female (28.7%).

3.1.3. Job role

Table 10. Job role

	Counts	Total	Proportion
Developer	92	150	0.613
Product/Project	26	150	0.173
Analyst	19	150	0.127
QA	5	150	0.033
UX Researcher	5	150	0.033
Product Designer	1	150	0.007
Sales	1	150	0.007
System and network administrator	1	150	0.007

Source: dataset analysed by author in JASP

The participants' job roles are mainly developers (61.3%), followed by product/project roles (17.3%), analysts (12.7%), and others with smaller proportions.

3.1.4. Grade

Table 11. Grade

	Counts	Total	Proportion
Intern/Junior	45	150	0.300
Middle	61	150	0.407
Senior	44	150	0.293

Source: dataset analysed by author in JASP

Participants are distributed across different grades, with Middle-level professionals (40.7%) being the most common, followed by Junior/Intern (30%), and Senior (29.3%). Their professional experience varies between 0 to 43 years, with an average of 5.4 years.

3.1.5. Company type

Table 12. Company type

	Counts	Total	Proportion
freelance	18	149	0.121
micro: 1-10 employees	10	149	0.067
small: 11-50 employees	26	149	0.174
medium: 51-500 employees	33	149	0.221
large: 501-5.000 employees	29	149	0.195
enterprise: 5.001-50.000 employees	20	149	0.134
giant: over 50.000 employees	13	149	0.087

Source: dataset analysed by author in JASP

The sample consists of professionals from diverse company sizes, with the majority being freelance (12.1%) and micro to medium-sized companies (6.7% to 22.1%).

3.1.6. Main country of company's operations

Table 13. Main country of company's operations

	Counts	Total	Proportion
Russia	66	150	0.440
Italy	33	150	0.220
Germany	16	150	0.107
USA	10	150	0.067
International	6	150	0.040
France	3	150	0.020
Israel	2	150	0.013
Kazakhstan	2	150	0.013
Armenia	1	150	0.007
Canada	1	150	0.007
Czech	1	150	0.007
English speaking country	1	150	0.007
Singapore	1	150	0.007
Spain	1	150	0.007
Switzerland	1	150	0.007
Syria	1	150	0.007
UK	1	150	0.007
Ukraine	1	150	0.007
Uzbekistan	1	150	0.007
Poland	1	150	0.007

Source: dataset analysed by author in JASP

The participants represent various countries, with the highest proportions from Russia (44%) and Italy (22%). We tried to keep in this way as the research is made under Russian and Italian institutions.

3.1.7. Main industry

Table 14. Main industry (of a company or of most of the clients, if it is a freelancer)

	Counts	Total	Proportion
Business/IT services (including Web and Mobile App	62	150	0.414
Banking and Finance	25	150	0.167
Retail and eCommerce	10	150	0.067
Telecom	8	150	0.053
Automotive	6	150	0.04
I am a freelancer who works in different industries	6	150	0.04
Travel, Hospitality, and Tourism	4	150	0.027

Education	3	150	0.02
Gamedev	3	150	0.02
Manufacturing	3	150	0.02
Cyber security	2	150	0.013
Healthcare and Pharma	2	150	0.013
Insurance IT	2	150	0.013
Logistics	2	150	0.013
Advertising	1	150	0.007
Architecture & Construction	1	150	0.007
Consulting	1	150	0.007
Crowdfunding	1	150	0.007
Energy and gas	1	150	0.007
Entertainment	1	150	0.007
Field Management Service	1	150	0.007
Highways infrastructures	1	150	0.007
Multimedia	1	150	0.007
Music	1	150	0.007
Non profit (open source)	1	150	0.007
SaaS	1	150	0.007

Source: dataset analysed by author in JASP

The participants work in various industries, with the most represented being Business/IT services (40.7%) and Banking and Finance (16.7%).

3.1.8. Frequency of usage

Table 15. How frequently do you use ChatGPT for work? (considering, working week = 5 days)

Level	Counts	Total	Proportion
1-2 days a week	21	146	0.144
2 or less days a week	3	146	0.021
3-4 days a week	30	146	0.205
Every day	29	146	0.199
Several times a month	38	146	0.260
Stopped using it at all	25	146	0.171

Source: dataset analysed by author in JASP

Participants reported different frequencies of ChatGPT usage, ranging from 1-2 days a week to every day.

3.1.9. ChatGPT version

Table 16. ChatGPT version

Level	Counts	Total	Proportion
3.5 (free)	120	149	0.805
4.0 (paid)	29	149	0.195

Source: dataset analysed by author in JASP

The majority of respondents (80.5%) use ChatGPT version 3.5 (free), while the rest (19.5%) use version 4.0 (paid).

3.1.10. Does the line manager aware of usage?

Table 17. Do(es) your line manager or your clients (if you are a freelancer) know(s) that you are using ChatGPT for work?

Level	Counts	Total	Proportion
Don't want to answer on this question	18	150	0.20
Just no	50	150	0.333
No, and it's even restricted	7	150	0.047
Yes, and he is promoting it's usage	6	150	0.040
Yes, and manager (clients) is (are) promoting it's usage	14	150	0.093
Yes, but it's my own initiative	55	150	0.367

Source: dataset analysed by author in JASP

An interesting fact, we also asked if the manager or a client side (in case if the respondent is a freelancer) know that the respondent is using ChatGPT for work. The results are saying that most respondents (36.7%) use ChatGPT with their manager's knowledge on their own initiative, while 38% use it without managerial knowledge or even against rules. Only 13.3% work in places where ChatGPT's usage is actively promoted. Around 12% chose not to disclose their manager's awareness status.

3.3 Description statistics of dependent variables

To understand how ChatGPT affects, several aspects of its influence on productivity and specific stages of software development were evaluated that were used in the analysis later as dependent variables. Appendix 2 is describing all the variables presented in dataset, the descriptions will not be repeated in this section. Here the focus will go to the values that were collected.

Table 18. Dimensions of productivity impact dependent variables: description statistics

	Overall_impact_productivity	Speed	More_tasks_same_time	Focus_higher_ lvl	Less_tired	Quality_improvement	Errors_reduce	Boost_problem_solving
Valid	150	150	150	150	150	150	150	150
Mean	2,8	3,2	2,3	2,7	2,5	2,6	2,2	2,6
Std, Deviation	1,3	1,4	1,4	1,4	1,4	1,4	1,3	1,4

Source: dataset analysed by author in JASP
Overall Impact on Performance: The average rating is 2.80 on a scale of 1 to 5. This indicates a moderate positive impact with a standard deviation of 1.253. Responses ranged from a minimum of 1 (indicating no impact) to a maximum of 5 (indicating extreme impact).
Speed: When it comes to completing tasks quickly, respondents view it positively. The average response was 3.24, suggesting that a significant number of respondents believe that ChatGPT speeds up their work.
Task management. The statement that ChatGPT helps with multitasking received an average rating of 2.34. This suggests that while some find it helpful, others may not see a clear difference.
Focus on higher level tasks. The average score of 2.74 indicates that ChatGPT helps developers focus on more strategic, higher-level tasks.
Reduced fatigue. A mean score of 2.48 indicates a moderate opinion that using ChatGPT can make completing tasks less tiring.
Quality of work: The mean value of 2.56 suggests a moderate opinion of ChatGPT improving the quality of work.
Reduce errors. The mean value of 2.207 indicates that somehow respondents believe that ChatGPT helps minimize errors in their results.
Problem Solving: The average score of 2.60 supports the view that ChatGPT can play a role in improving the problem-solving skills.

Notably, there are missing values (Table 19) due to the varied involvement of respondents in different software development stages.

Table 19. Influence on specific software development stages: description statistics

	Requirements_1	Software_design_1	Coding_1	Code_review_1	Testing_QA_1	Deployment_1	Maintenance_bug_fix_1	Documentation_1
Valid	67	66	109	61	50	41	76	79
Missing	83	84	41	89	100	109	74	71
Mean	2,2	2,2	2,7	2,5	2,3	1,8	2,4	2,8
Std, Deviation	1,1	1,0	1,1	1,4	1,3	1,0	1,2	1,3

Source: dataset analysed by author in JASP
Planning: With 69 valid responses and 81 missing values, the planning stage yielded a mean score of 2.406. This suggests a moderate time-saving potential of ChatGPT in this phase.
Requirements: Out of the 67 valid responses, the requirements gathering phase received a mean score of 2.179.
Software Design: A mean score of 2.167 from 66 respondents' hints at a modest time-saving benefit of ChatGPT during the design phase.

Coding: Given that coding is a fundamental aspect of software development, it's unsurprising that it had 109 valid responses. The mean score of 2.688 suggests a favourable time-saving potential of ChatGPT.
Code Review: 61 valid responses produced a mean score of 2.475, which implies a moderate perception of time saved during the code review process.
Testing & QA: With 50 valid responses, testing and quality assurance phases recorded a mean score of 2.320, showcasing a somehow positive time-saving sentiment.
Deployment: Only 41 developers responded regarding the deployment stage, yielding a lower average rating of 1.756. This may suggest either limited involvement in this stage or a perception that ChatGPT has lesser utility here.
Maintenance and Bug Fixes: The 76 valid responses resulted in a mean score of 2.368, pointing towards a positive sentiment about ChatGPT's efficiency in this domain.
Documentation: 79 responses indicated a higher mean score of 2.848. This highlights that respondents might find ChatGPT particularly beneficial for documentation tasks.

3.3. Correlations for examination of relationships between perception of general ChatGPT's influence and different respondent groups

After having the first glance, correlations related to overall feedback on ChatGPT's influence on productivity were identified. Were studied all the possible variants of groups. For making correlation analysis, was used Kendall's coefficient, as the data are not normally distributed and they are also mostly representing ordinal values.

3.3.1.Grade

Table 20. Kendall's tau correlations for grade

Variable		Grade	Strength
1. Grade	Kendall's Tau B	—	
	p-value	—	
2. Overall_impact_productivity	Kendall's Tau B	-0.129	
	p-value	0.064	
3. Speed	Kendall's Tau B	-0.14	weak
	p-value	0.043	
4. More_tasks_same_time	Kendall's Tau B	-0.129	
	p-value	0.068	
5. Focus_higher_lvl	Kendall's Tau B	-0.073	
	p-value	0.291	
6. Less_tired	Kendall's Tau B	-0.179	weak
	p-value	0.01	
7. Quality_improvement	Kendall's Tau B	-0.199	weak to moderate
	p-value	0.004	

8. Errors_reduce	Kendall's Tau B	-0.154	weak
	p-value	0.029	
9. Boost_problem_solving	Kendall's Tau B	-0.207	weak to moderate
	p-value	0.003	

Source: dataset analysed by author in JASP

The adverse According to Kendall's Tau B values, there is an inverse correlation between Grade and the factors we looked at. It suggests that the apparent benefits at this moment decrease as the grade level rises (from Intern/Junior to Senior).
Let's examine the noteworthy outcomes:

- **SPEED:** $\tau = -0.14$, $p = 0.043$. It means, Grade increases, the perception of ChatGPT helping to complete tasks more quickly slightly decreases.
- **LESS_TIRED:** $\tau = -0.179$, $p = 0.01$. Grade increases -> perception decreases.
- **QUALITY_IMPROVEMENT:** $\tau = -0.199$, $p = 0.004$. Senior-level staff are less likely to perceive ChatGPT as improving the quality of their work.
- **ERRORS_REDUCE:** $\tau = -0.154$, $p = 0.029$. Seniors less likely will thing that ChatGPT can reduce errors in their work.
- **BOOST_PROBLEM_SOLVING:** $\tau = -0.207$, $p = 0.003$. Seniors don't think that ChatGPT enhances their problem-solving abilities.

3.3.2.Experience

Table 21. Kendall's tau correlations for experience

Variable		Experience	Strength
1. Experience	Kendall's Tau B	—	
	p-value	—	
2. Overall_impact_productivity	Kendall's Tau B	-0.138	weak
	p-value	0.028	
3. Speed	Kendall's Tau B	-0.177	weak
	p-value	0.005	
4. More_tasks_same_time	Kendall's Tau B	-0.162	weak
	p-value	0.011	
5. Focus_higher_lvl	Kendall's Tau B	-0.127	
	p-value	0.044	
6. Less_tired	Kendall's Tau B	-0.217	weak to moderate
	p-value	< .001	
7. Quality_improvement	Kendall's Tau B	-0.233	moderate
	p-value	< .001	
8. Errors_reduce	Kendall's Tau B	-0.194	weak to moderate
	p-value	0.002	
9. Boost_problem_solving	Kendall's Tau B	-0.129	
	p-value	0.041	

Source: dataset analysed by author in JASP

Similar to Grade, negative τ values means that when Experience increases, perceived benefits of ChatGPT at this time decrease.
Again, let's look at significant results:

- **SPEED:** $\tau = -0.177$, $p = 0.005$. As Experience increases, the perception of ChatGPT helping to complete tasks more quickly decreases.
- **MORE_TASKS_SAME_TIME:** $\tau = -0.162$, $p = 0.011$. More experienced professionals will less likely feel that ChatGPT helps them perform more tasks at the same time.
- **LESS_TIRED:** $\tau = -0.217$, $p < 0.001$. More experienced respondents say that they are less likely feel less tired if they complete tasks with ChatGPT.
- **QUALITY_IMPROVEMENT:** $\tau = -0.233$, $p < 0.001$. The more experienced the respondent is, the less likely he will think ChatGPT can improve the quality of his work.
- **ERRORS_REDUCE:** $\tau = -0.194$, $p = 0.002$. Same with the reducing errors in work.

In overall, Grade and Experience both say that younger workers perceive ChatGPT as more beneficial.

3.3.3. Company size

Table 22. Kendall's tau correlations for Company size

Variable		Company_size	Strength
1. Company_size	Kendall's Tau B	—	
	p-value	—	
2. Overall_impact_productivity	Kendall's Tau B	-0.147	
	p-value	0.017	
3. Speed	Kendall's Tau B	-0.168	weak
	p-value	0.006	
4. More_tasks_same_time	Kendall's Tau B	-0.221	weak to moderate
	p-value	< .001	
5. Focus_higher_lvl	Kendall's Tau B	-0.231	moderate
	p-value	< .001	
6. Less_tired	Kendall's Tau B	-0.198	weak to moderate
	p-value	0.001	
7. Quality_improvement	Kendall's Tau B	-0.203	weak to moderate
	p-value	0.001	
8. Errors_reduce	Kendall's Tau B	-0.139	
	p-value	0.027	
9. Boost_problem_solving	Kendall's Tau B	-0.082	
	p-value	0.185	

Source: dataset analysed by author in JASP

This is the surprise, as we did not think about possible correlation while setting the hypothesis, however, negative τ values show that as the Company size increases, the perceived benefits of ChatGPT decrease.
Larger companies seem to have employees who perceive fewer benefits from ChatGPT. This could be explained by the fact that processes tools, workflows might be more established.

3.3.4. ChatGPT version

Table 23. Kendall's tau correlations for ChatGPT version

Variable		**ChatGPT_version**	**Strength**
1. ChatGPT_version	Kendall's Tau B	—	
	p-value	—	
2. Overall_impact_productivity	Kendall's Tau B	0.25	moderate
	p-value	< .001	
3. Speed	Kendall's Tau B	0.162	weak
	p-value	0.027	
4. More_tasks_same_time	Kendall's Tau B	0.259	moderate
	p-value	< .001	
5. Focus_higher_lvl	Kendall's Tau B	0.234	moderate
	p-value	0.002	
6. Less_tired	Kendall's Tau B	0.13	
	p-value	0.079	
7. Quality_improvement	Kendall's Tau B	0.176	weak
	p-value	0.018	
8. Errors_reduce	Kendall's Tau B	0.142	
	p-value	0.058	
9. Boost_problem_solving	Kendall's Tau B	0.176	weak
	p-value	0.018	

Source: dataset analysed by author in JASP

For ChatGPT version, users of the paid version (ChatGPT 4.0) generally perceive more benefits in productivity, speed, multitasking, focus on higher-level tasks, quality of work, and problem-solving.
The correlation analysis allowed as to identify relationships between overall perceptions of ChatGPT's influence and such groups as Grade, Age, Experience, ChatGPT version, Company size.
Saying in short, as respondents' professional grade or experience increases, they tend to perceive less benefit from using ChatGPT across various areas, including speed, quality improvement, and problem-solving. Employees from larger companies perceive fewer benefits from ChatGPT, what might be related to the fact that in larger organizations processes are more established.
Users of the newer or paid version of ChatGPT (presumably ChatGPT 4.0) tend to report more benefits in areas such as productivity and multitasking.

However, these correlations, while statistically significant, are mostly weak and only sometimes moderate in strength. It means there is a relationship but it is not necessarily that these are the dominant effects.

3.4. ANOVA test for finding differences between groups

After understanding the overall perception of ChatGPT's influence on different aspects of work, it is time to analyze the differences between different groups of respondents.
To do this, we used an ANOVA test as it allows to compare the means of 3 or more groups to understand is there is a statistically significant difference between them.

3.4.1. Influence on Overall productivity

3.4.1.1. Influence of Grade

Table 24. ANOVA - Overall_impact_productivity

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Grade	6.435	2	3.218	2.078	0.129	0.028
Residuals	227.565	147	1.548			

Source: dataset analysed by author in JASP

Table 25. Descriptives - Overall_impact_productivity

Grade	N	Mean	SD	SE	Coefficient of variation
Intern/Junior	45	3.111	1.301	0.194	0.418
Middle	61	2.705	1.256	0.161	0.464
Senior	44	2.614	1.166	0.176	0.446

Source: dataset analysed by author in JASP

Looking at the descriptives, we see that Intern/Junior employees have a slightly higher mean perceived productivity (3.111) than Middle (2.705) and Senior employees (2.614). However, the difference isn't statistically significant.

3.4.1.2. Influence of Age group

Table 26. ANOVA - Overall_impact_productivity

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Age_group	13.19	3	4.397	2.907	0.037	0.056
Residuals	220.81	146	1.512			

Source: dataset analysed by author in JASP

Table 27. Descriptives - Overall_impact_productivity

Age_group	N	Mean	SD	SE	Coefficient of variation
18-24	36	3	1.242	0.207	0.414
25-34	84	2.869	1.2	0.131	0.418
35-44	28	2.25	1.295	0.245	0.575
45 and above	2	4	1.414	1	0.354

Source: dataset analysed by author in JASP

The F-statistic of 2.907 with a p-value of 0.037 indicates that there is a statistically significant difference in perceived productivity among the different age groups. The age group of 45 and above reported the highest mean perceived productivity (4.0), followed by the 18-24 age group (3.0), then the 25-34 age group (2.869), and lastly the 35-44 age group (2.25). Considering that in group 45+ we have only 2 participants, we don't count this result as significant. For the rest of the groups H1 that younger groups perceive more benefits is proved.

3.4.1.3. Influence of Company type by size

Table 28. ANOVA - Overall_impact_productivity

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Company_type	14.076	6	2.346	1.549	0.166	0.061
Residuals	215.052	142	1.514			

Source: dataset analysed by author in JASP

Table 29. Descriptives - Overall_impact_productivity

Company_type	N	Mean	SD	SE	Coefficient of variation
enterprise: 5.001-50.000 employees	20	2.45	1.146	0.256	0.468
freelance	18	3.444	1.381	0.326	0.401
giant: over 50.000 employees	13	2.538	1.127	0.312	0.444
large: 501-5.000 employees	29	2.517	1.379	0.256	0.548

medium: 51-500 employees	33	2.97	1.159	0.202	0.39
micro: 1-10 employees	10	2.8	0.789	0.249	0.282
small: 11-50 employees	26	2.769	1.275	0.25	0.46

Source: dataset analysed by author in JASP

Freelancers reported the highest mean perceived productivity (3.444), while large-sized companies (501-5,000 employees) had the lowest (2.517). Differences are, however, not statistically significant with an F-statistic of 1.549 and a p-value of 0.166.

3.4.1.4. Influence of Company type by ChatGPT version

Table 30. ANOVA - Overall_impact_productivity

Cases	Sum of Squares	df	Mean Square	F	p	η²
ChatGPT_version	19.294	1	19.294	13.517	< .001	0.084
Residuals	209.833	147	1.427			

Source: dataset analysed by author in JASP

Table 31. Descriptives - Overall_impact_productivity

ChatGPT_version	N	Mean	SD	SE	Coefficient of variation
3.5 (free)	120	2.608	1.176	0.107	0.451
4.0 (paid)	29	3.517	1.271	0.236	0.361

Source: dataset analysed by author in JASP

The F-statistic of 13.517 with a p-value of < .001 indicates a statistically significant difference in perceived productivity between users of different ChatGPT versions.
Users of the paid version (ChatGPT 4.0) reported higher perceived productivity (3.517) than users of the free version (ChatGPT 3.5), which stood at 2.608. This could mean that those investing in the paid version find it more impactful in terms of productivity.

3.4.2. Influence of Grade on Speed of completing tasks with ChatGPT

Table 32. ANOVA – Speed

Cases	Sum of Squares	df	Mean Square	F	p	η²
Grade	9.608	2	4.804	2.506	0.085	0.033
Residuals	281.752	147	1.917			

Source: dataset analysed by author in JASP

Table 33. Descriptives – Speed

Grade	N	Mean	SD	SE	Coefficient of variation
Intern/Junior	45	3.622	1.336	0.199	0.369
Middle	61	3.115	1.33	0.17	0.427
Senior	44	3.023	1.502	0.226	0.497

Source: dataset analysed by author in JASP

The difference in perceived speed across grades may be approaching statistical significance, but not at the 0.05 level, according to the F-statistic of 2.506 and p-value of 0.085. The highest mean speed was recorded by intern/junior employees (3.622), followed by middle (3.115) and senior (3.023) personnel. This would imply that less seasoned or younger employees find ChatGPT to be more responsive or useful, however this trend is not very statistically significant.

3.4.3. Influence of Grade on possibility to perform several tasks for the same time with ChatGPT

Table 34. ANOVA - More_tasks_same_time

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Grade	9.492	2	4.746	2.641	0.075	0.035
Residuals	264.168	147	1.797			

Source: dataset analysed by author in JASP

Table 35. Descriptives - More_tasks_same_time

Grade	N	Mean	SD	SE	Coefficient of variation
Intern/Junior	45	2.711	1.29	0.192	0.476
Middle	61	2.115	1.279	0.164	0.605
Senior	44	2.273	1.468	0.221	0.646

Source: dataset analysed by author in JASP

The ability to do multiple activities at once across different grades is approaching statistical significance, but is not significant at the 0.05 level, according to the F-statistic of 2.641 and p-value of 0.075.
The highest ability to manage several tasks at once was indicated by intern/junior employees (mean = 2.711), followed by senior employees (mean = 2.273) and middle-grade employees (mean = 2.115). Although the observed trend isn't significantly statistically significant, it shows that younger or entry-level employees may find tools like ChatGPT more effective for multitasking.

3.4.4. Influence of Grade on feeling less tired with ChatGPT

Table 36. ANOVA - Less_tired

Cases	Sum of Squares	df	Mean Square	F	p	η²
Grade	22.273	2	11.137	5.993	0.003	0.075
Residuals	273.167	147	1.858			

Source: dataset analysed by author in JASP

Table 37. Descriptives - Less_tired

Grade	N	Mean	SD	SE	Coefficient of variation
Intern/Junior	45	3.067	1.514	0.226	0.494
Middle	61	2.197	1.249	0.16	0.569
Senior	44	2.273	1.353	0.204	0.595

Source: dataset analysed by author in JASP

There is a statistically significant difference in the felt reduction of weariness across different grades, according to the F-statistic of 5.993 and p-value of 0.003.
Intern/Junior employees feel the least tired (mean = 3.067) when using ChatGPT, indicating that they might find the tool more alleviating in terms of workload or cognitive demands. This is in contrast to Middle (mean = 2.197) and Senior employees (mean = 2.273), who reported feeling more tired. It might mean that younger or less experienced employees perceive a greater benefit in workload reduction from ChatGPT than their senior counterparts.

3.4.5. Influence of Grade on Quality improvement of work with ChatGPT

Table 38. ANOVA - Quality_improvement

Cases	Sum of Squares	df	Mean Square	F	p	η²
Grade	18.834	2	9.417	5.405	0.005	0.068
Residuals	256.126	147	1.742			

Source: dataset analysed by author in JASP

Table 39. Descriptives - Quality_improvement

Grade	N	Mean	SD	SE	Coefficient of variation
Intern/Junior	45	3.089	1.379	0.206	0.446
Middle	61	2.41	1.371	0.176	0.569
Senior	44	2.227	1.179	0.178	0.529

Source: dataset analysed by author in JASP

With a p-value of 0.005 and an F-statistic of 5.405, which is statistically significant at the 0.05 level, it can be seen that there are disparities in how the perceived quality of improvement is felt to have progressed among the grades.
The biggest perceived improvement in quality is perceived by intern/junior employees (mean = 3.089), followed by senior employees (mean = 2.227), and middle-grade employees (mean = 2.41).
This implies younger or less experienced employees feel that ChatGPT has a more positive impact on the quality of their work compared to their senior counterparts.

3.4.6. Influence of Grade on errors reduce with ChatGPT

Table 40. ANOVA - Errors_reduce

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Grade	8.901	2	4.45	2.776	0.066	0.036
Residuals	235.693	147	1.603			

Source: dataset analysed by author in JASP

Table 41. Descriptives - Errors_reduce

Grade	N	Mean	SD	SE	Coefficient of variation
Intern/Junior	45	2.578	1.305	0.195	0.506
Middle	61	2.066	1.237	0.158	0.599
Senior	44	2.023	1.267	0.191	0.626

Source: dataset analysed by author in JASP

The F-statistic is 2.776 with a p-value of 0.066. This is approaching significance at the 0.05 level but is not there yet, suggesting potential differences in error reduction across grades but not strong enough to be statistically significant.
Intern/Junior employees report the highest reduction in errors (mean = 2.578) compared to Middle (mean = 2.066) and Senior employees (mean = 2.023). Again, younger employees seem to find a greater reduction in errors using ChatGPT, though this trend isn't strongly statistically supported.

3.4.7. Influence of Grade on possibility to boost problem solving skills with ChatGPT

Table 42. ANOVA - Boost_problem_solving

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Grade	17.605	2	8.803	4.894	0.009	0.062
Residuals	264.395	147	1.799			

Source: dataset analysed by author in JASP

Table 43. Descriptives - Boost_problem_solving

Grade	N	Mean	SD	SE	Coefficient of variation
Intern/Junior	45	3.067	1.498	0.223	0.489
Middle	61	2.557	1.323	0.169	0.517
Senior	44	2.182	1.187	0.179	0.544

Source: dataset analysed by author in JASP

The F-statistic of 4.894 with a p-value of 0.009 indicates a statistically significant difference in the perceived boost in problem-solving across different grades. Intern/Junior employees feel a stronger boost in problem-solving (mean = 3.067), in contrast to Middle (mean = 2.557) and Senior employees (mean = 2.182). Younger employees perceive a greater aid in problem-solving from ChatGPT compared to more experienced individuals.

3.4.8. Influence of Grade on possibility to focus with ChatGPT

Table 44. ANOVA - Focus_higher_lvl

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Grade	13.116	2	6.558	3.471	0.034	0.045
Residuals	277.744	147	1.889			

Source: dataset analysed by author in JASP

Table 45. Descriptives - Focus_higher_lvl

Grade	N	Mean	SD	SE	Coefficient of variation
Intern/Junior	45	3.111	1.385	0.207	0.445
Middle	61	2.41	1.321	0.169	0.548
Senior	44	2.818	1.435	0.216	0.509

Source: dataset analysed by author in JASP

The statistically significant F-statistic of 3.471 and p-value of 0.034 point to differences in students' capacity to focus at a greater level across grades. Senior employees are in the middle (mean = 2.818), while Middle-grade employees reported the least (mean = 2.41), and Intern/Junior employees reported the strongest capacity to focus at a higher level (mean = 3.111). In terms of improved focus, ChatGPT seems to help younger employees more.

3.5. Means test for ChatGPT's influence on different stages analysis

After understanding the differences in groups of overall perception of ChatGPT's influence on different aspects of work, it is time to analyze the answers related to our main goal of research – to different stages of software development.
To do this, we used Means, Medians, Modes analysis as it suits our goals best.
Table 46. Stages of software development and ChatGPT's influence: overall picture

	Plan	Requir	Soft design	Coding	Code review	QA	Deploy-ment	Maintenance Bug fix	Documen-tation
Valid	69	67	66	109	61	50	41	76	79
Mode	1	2	2	3	1	1	1	2	3
Median	2	2	2	3	2	2	1	2	3
Mean	2.406	2.179	2.167	2.688	2.475	2.32	1.756	2.368	2.848
Std. Deviation	1.204	1.072	1.046	1.111	1.361	1.253	1.019	1.164	1.282
Minimum	1	1	1	1	1	1	1	1	1
Maximum	5	5	5	5	5	5	5	5	5

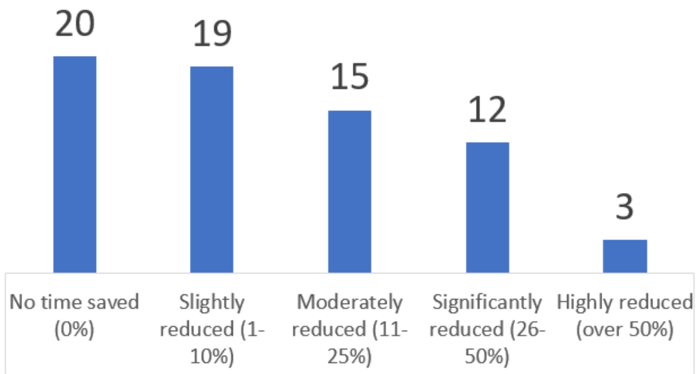
Source: dataset analysed by author in JASP

Respondents distribution:

- Coding has the highest number of respondents (109), but this can be explained by the fact that most of our respondents are Developers.
- Deployment has the fewest respondents (41). Considering that in this stage developers are also involved, it probably means that fewer developers or teams have found that ChatGPT can be used for this stage.

3.5.1. Planning

Figure 9. Answers distribution for Planning



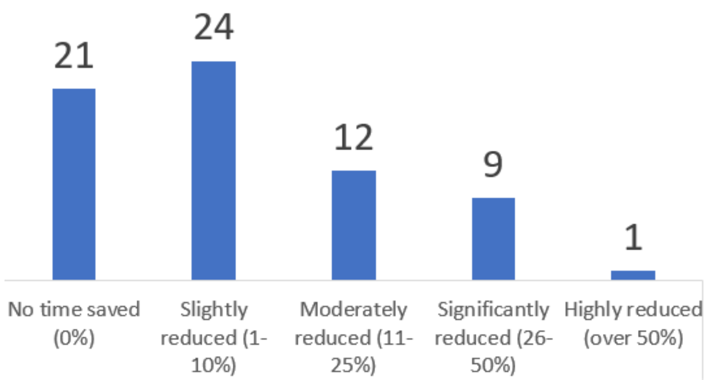
Source: made by author on a basis of gathered data

- **Validation (Valid)**: 69 respondents answered for this stage.
- **Mean Rating**: 2.406, on average. It is in the middle between slightly and moderately reduces.
- **Mode**: 1, which means that most respondents feel no time is saved in this stage by using ChatGPT.
- **Median**: 2, suggesting that the middle value is a slight reduction in time.

Given that planning is a high-level, conceptual phase, it's interesting that respondents find value in using AI here. They might be using ChatGPT for brainstorming, getting clarity on certain topics.

3.5.2. Requirements

Figure 10. Answers distribution for Requirements



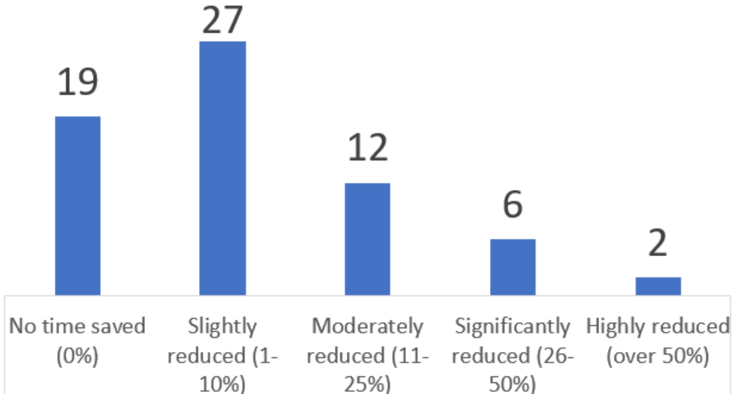
Source: made by author on a basis of gathered data

- **Valid**: 67 respondents answered for this stage.
- **Mean Rating**: 2.179, suggesting a slight reduction in time.
- **Mode**: 2, meaning most users feel there's a slight reduction in time.
- **Median**: 2, indicating a slight reduction in time.

ChatGPT might assist in better understanding of requirements, especially when facing with unfamiliar terminologies or when there is a need to explain technical details in these terms.

3.5.3. Software Design

Figure 11. Answers distribution for Software Design



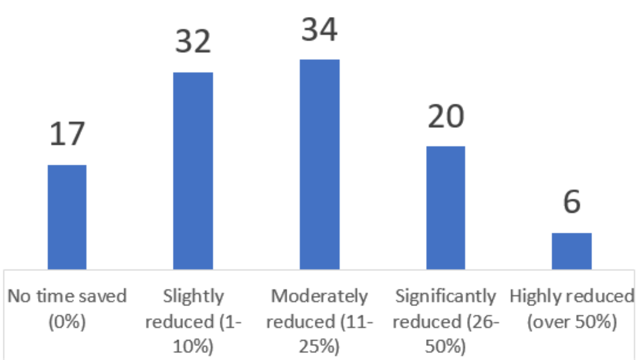
Source: made by author on a basis of gathered data

- **Valid**: 66 respondents.
- **Mean Rating**: 2.167, indicating a slight reduction in time.
- **Mode**: 2, most users perceive a slight time-saving.
- **Median**: 2, reinforcing the slight reduction.

This phase involves a lot of conceptual work. The consistent ratings suggest respondents might be using ChatGPT to understand best practices or get insights into specific design patterns and architectures. However, the methods of usage of ChatGPT in this stage might be studied additionally for a better understanding.

3.5.4. Coding

Figure 12. Answers distribution for Coding



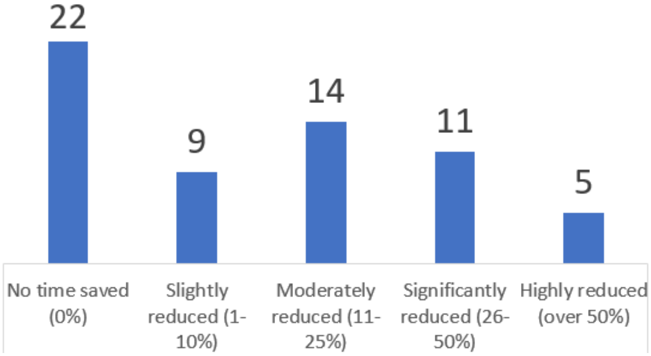
Source: made by author on a basis of gathered data

- **Valid**: 109 respondents.
- **Mean Rating**: 2.688, suggesting that users believe ChatGPT moderately reduces time during coding.
- **Mode**: 3, indicating that the most common response is a moderate reduction.
- **Median**: 3, reinforcing the moderate reduction in time.

As the stage with the highest number of respondents and a higher average rating, it's evident that developers find value in ChatGPT when actively writing code. They might be using it to troubleshoot errors, understand specific functions, or seek coding best practices.

3.5.5. Code review

Figure 13. Answers distribution for Code review



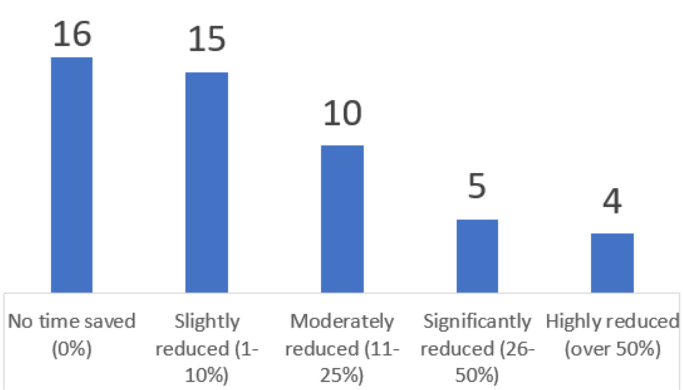
Source: made by author on a basis of gathered data

- **Valid**: 61 respondents.
- **Mean Rating**: 2.475, between slight and moderate reduction.
- **Mode**: 1, which suggests many respondents feel no time-saving in this stage.
- **Median**: 2, indicating a slight time-saving.

This stage is intriguing. While some respondents see no time savings, others find it beneficial. ChatGPT might be assisting in understanding certain code structures or logic. However, the varied ratings also hint that a code review's qualitative nature might not always align with ChatGPT's capabilities.

3.5.6. Testing and QA

Figure 14. Answers distribution for Testing and QA



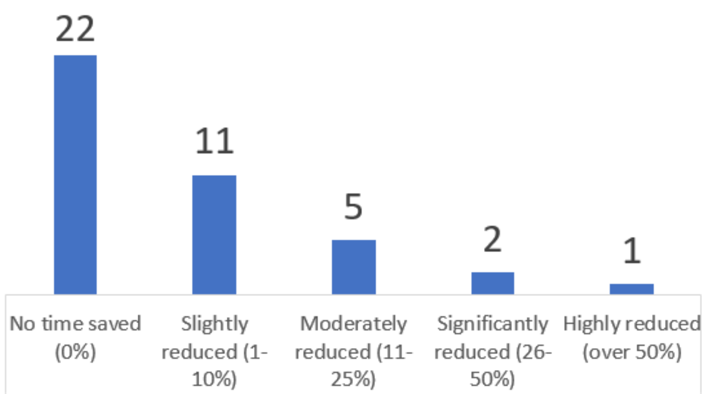
Source: made by author on a basis of gathered data

- **Valid**: 50 respondents.
- **Mean Rating**: 2.32, suggesting a slight to moderate reduction in time.
- **Mode**: 1, with many feeling no time-saving.
- **Median**: 2, reinforcing the slight time-saving.

ChatGPT could be assisting QA engineers in understanding certain bugs, forming test scenarios, or perhaps clarifying how specific features should work based on descriptions. However, we would like to note that the most common answer is no time saved. This is reasonable, because ChatGPT is not capable of crawling on webpages or in apps so QA specialists hardly can implement ChatGPT for bug searches.

3.5.7. Deployment

Figure 15. Answers distribution for Deployment



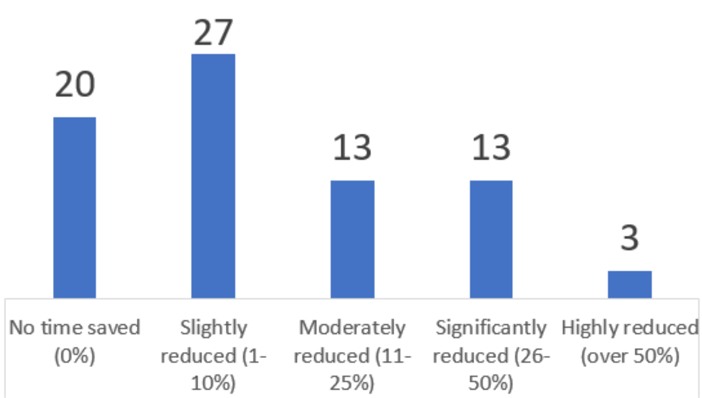
Source: made by author on a basis of gathered data

- **Valid**: 41 respondents.
- **Mean Rating**: 1.756, closer to no time-saving.
- **Mode**: 1, many respondents see no benefit here.
- **Median**: 1, reinforcing the notion of no time-saving.

ChatGPT is not helping much in Deployment. Users might be consulting it for deployment best practices, troubleshooting, or understanding deployment logs, but given that deployment can often be a scripted or automated process, ChatGPT is not capable of doing the process by itself.

3.5.8. Maintenance and Bug fix

Figure 16. Answers distribution for Maintenance and Bug fix



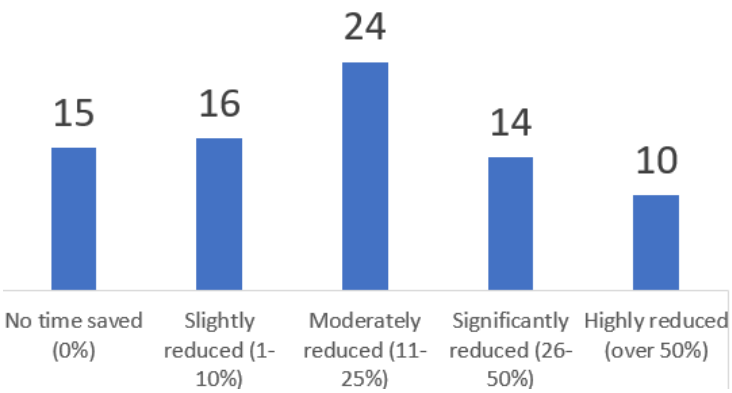
Source: made by author on a basis of gathered data

- **Valid**: 76 respondents.
- **Mean Rating**: 2.368, suggesting a slight to moderate reduction in time.
- **Mode**: 2, most users see a slight time-saving.
- **Median**: 2, further supporting the slight time-saving perception.

Maintenance involves bug fixing, code refactoring, and updates. ChatGPT is slightly supporting respondents in this stage. Probably its assistance can be useful in understanding legacy code, finding solutions to recurring issues, or exploring best practices for code optimization.

3.5.9. Documentation

Figure 17. Answers distribution for Documentation



Source: made by author on a basis of gathered data

- **Valid**: 79 respondents.
- **Mean Rating**: 2.848, suggesting a moderate time-saving when using ChatGPT for documenting bug fixes.
- **Mode**: 3, indicating a moderate time-saving.
- **Median**: 3, reinforcing the moderate reduction in time.

Documenting is crucial for team knowledge. Respondents say that ChatGPT is moderately helpful during this stage. This is reasonable, as ChatGPT, being a language model, is trained to create professional texts. ChatGPT can suggest documentation structure, or even write it, if a professional will make a prompt with idea.

Coming to the conclusion of this part**, Coding** and **Documentation** stages show the highest perceived time-saving when using ChatGPT.

The least apparent benefit of ChatGPT appears to be **in deployment. Deployment and QA, however,** had far lower mean scores and fewer respondents. When it comes to deployment, predetermined procedures and scripts are frequently used, therefore ChatGPT might not be needed much. This could indicate to QA that either ChatGPT doesn't work properly in this stage or that users are unaware of how to use it well. In addition, since developers make up the majority of our respondents, it is possible that QA is not active in this stage. In order to comprehend the reasoning better, more study is required.

The second stage in the rating of those that benefit from ChatGPT less is **Software Design**. Its mean rating is 2.167. This could indicate that while ChatGPT provides value in many stages, the intuitive and visionary process of software design might be something where human expertise and creativity dominate.

Interesting to note that in stages **Code Review** and **QA**, the mode is 1 (No time saved), but their mean values are closer to 2.5. This indicates a polarization in user feedback: a significant number of users don't find any time-saving, while others might find considerable benefits.

3.6. Means test for ChatGPT's influence on stages analysis: what benefits most: creativity or a standard

One of our hypotheses suggests that ChatGPT may be more beneficial for less creative stages (H6. The impact of ChatGPT on software development productivity is more significant in less creative stages).
We asked our respondents to choose from the list of 9 stages 4 or less they consider the most creative. It allowed us to combine votes and to make a ranking list.
Table 47. Ranking of stages (from the 'most creative' to the 'least)

Stage	Votes	Result
Software design	83	More creative
Coding	78	More creative
Planning	50	More creative
Documentation	46	More creative
Requirement analysis	41	Less creative
Code review	24	Less creative
Maintenance and bug fixing	22	Less creative
Testing and QA	17	Less creative
Deployment	11	Less creative

Source: made by author on a basis of gathered data
This ranking allowed us to group stages

- **More creative Stages**: Software design, Coding, Planning, and Documentation.
- **Less-Creative Stages**: Requirement analysis, Code review, Maintenance and bug fixing, Testing and QA, and Deployment

We calculated means of time saving for these 2 big groups.
Table 48. Descriptive Statistics

	Less creative	More creative
Valid	295	323
Mode	1	2
Median	2	2
Mean	2.254	2.56
Std. Deviation	1.198	1.184
Minimum	1	1
Maximum	5	5

Source: made by author on a basis of gathered data
More creative Tasks:

- **Valid**: 323 votes.

- **Mean Rating**: 2.56. This value suggests a moderate time-saving when using ChatGPT for creative tasks.
- **Mode**: 2. This is the most frequently occurring value, indicating that a slight time-saving (1-10%) is the most common experience among respondents.
- **Median**: 2. The middle value, when the ratings are arranged in ascending order, also suggests a slight time-saving.

For tasks considered more creative, the data says that professionals mostly experience from slight to moderate reduction in time when using ChatGPT. This suggests that while the tool might not replace the need for human creativity, it aids in streamlining the process and making it more efficient.

Less Creative Tasks:

- **Valid**: 295 responses.
- **Mean Rating**: 2.254. This suggests a slight to moderate time-saving when employing ChatGPT for tasks considered less creative.
- **Mode**: 1. It indicates that the most common experience is that there's no time saved when using ChatGPT for these tasks.
- **Median**: 2. This middle value suggests a slight time-saving, which means there's a split, with half of the respondents experiencing no or very little time saving and the other half experiencing more noticeable benefits.

The results are mixed for tasks that are more routine in their nature. While the median and mean are saying about the presence of considerable time savings, the mode is showing that a large number of respondents did not experience any time savings using ChatGPT. This may happen by the reason that the nature of these tasks is more standardized and less reliant on outside information or support.

The distinction between creative and non-creative tasks provides a fascinating insight into how respondents perceive the value of AI. While one might expect a more obvious difference in the utility of ChatGPT between more creative and less creative tasks, the findings suggest a more nuanced picture. Both groups are being influenced in the pretty same way.

3.7. Participants speaking

We left a line for comments and suggestions in our survey and were truly impressed when our participants left a lot of meaningful comments about ChatGPT usage experience. We decided to separate them into this section as we think that the comments provided by participants offer valuable insights.

Senior Product Manager, 32 y.o., Entertainment field, main country - USA: "I like using it to write better, come up with better language structures/words that I would use at work."

Junior Developer/Engineer, 22 y.o., Business/IT services, main country - Russia: "For me, ChatGPT usually works as an improved Google. It doesn't automate my work, it just helps me to find answers faster."

Regarding coding tasks, opinions are varied. Although most of the respondents claim that for programming purposes it is not very useful:

1) Middle Developer/Engineer, 28 y.o., Banking and Finance, main country – USA: " I don’t have a lot of experience using ChatGPT for coding, I used for other purposes, for example to explore vision and mission of other companies. For programming purposes, It can help you if you're a freelancer that creates mostly static websites and things like that, but for big projects that use multiple languages and have a bigger complexity, I think that it's not so useful at the moment.".

2) Senior Developer/Engineer, 36 y.o., Retail and eCommerce, main country – Russia: ‘I tried to solve several coding problems with it and the answers were incorrect…’.

3) Senior Developer/Engineer, 36 y.o., IT, main country – Russia ‘ChatGPT generates a solution that is not conceptually correct. The time spent rewriting the code is an order of magnitude greater than the savings from automatic code generation. So far, ChatGPT cannot replace even a jun. After three months of testing, we stopped trying to use ChatGPT for development. Let's repeat the exercise when GPT 6.0 appears’.

However, another participant (Intern, Freelancer, Developer/Engineer, 27 y.o., Business/IT services, main country – Italy) that marked ChatGPT's help in Coding stage as ‘significantly reduced’ stating: "The level of understanding and the quality of responses it provides are truly impressive. ChatGPT has not only saved me time and effort but also expanded my knowledge and helped me explore new perspectives. I highly recommend giving it a try! It's an invaluable resource that can enhance your productivity and provide a delightful and insightful experience.

Some respondents highlight limitations.

1) Middle Business / System Analyst, 23 y.o., Banking and Finance, main country – Russia: ‘I don't find a lot of use cases for ChatGPT bc my industry is narrow and it doesn't know technical landscape of my company, so it can give me general observations and ideas but not realization advice, which is the most struggling part of my job’.

2) Middle Business / System Analyst, 29 y.o., Logistics, main country – USA: ‘Used a little at the very beginning when it first became a hot topic - the ability to write code, format it, find errors and comment on it is very impressive, but for routine use it is not suitable for coding (because there are too many details that will take longer to describe, than to write the same code yourself) + sometimes makes mistakes, which then need to be looked for and corrected, and in general, I want to write many things myself, and even if ChatGPT can do it for me, I don’t want to’.

Two participants highlighted the problem of information privacy and breaches, saying that they are trying to not put corporate information into ChatGPT while asking questions.

1) Middle Developer/Engineer, 33 y.o., Insurance software, main country – USA: ‘I use ChatGPT only for things that doesn't require sharing details about work project and for things that I am able to do myself, but longer. Mostly it is pure technical coding tasks’.

2) Senior Product Manager, 32 y.o., Entertainment field, main country – USA: ‘I also try to minimize and not share any company information with it and it limits how much I can use ChatGPT as a Product manager’.

Interestingly, some respondents reflected on the broader implications of large language models like ChatGPT. They believe that these models could lead to a shift in the human-computer relationship and play a transformative role in the future of technology.

One participant, Junior Product or Project Manager, 27 y.o., Architecture & Construction, main country – Russia, shared, "I see big language models as a promising tool to help you get down to work and also as a means of increasing productivity. In a global perspective, I believe that man's supreme intellectual primacy will be ceded to intelligence based on another form. Because humans and their thinking are subject to a large number of negative factors when evaluated in terms of the impact of uncertainty, scalability, thinking errors, patterns, memory capacity, number of operations per second, fault tolerance, etc. Yes, this is a dehumanising assessment, but the progress of technology is not prone to pity.

Participants note that ChatGPT is undeniably good at helping with documentation. Senior Developer/Engineer, 36 y.o., Retail and eCommerce, main country – Russia: ‘ChatGPT can be useful in the text transformation field, make it shorter/longer, express the meaning in simple or complex way’.

Overall, the comments demonstrate what was illustrated in statistical interpretation of results: that ChatGPT has opportunities and limitations and they are related to different software development stages. Comments also says ChatGPT can be used at various stages as a search engine and additionally rase questions of data privacy.

Chapter 4. Discussion and Conclusion

This section will sum up the results that we obtained in the previous part. We will think on how our research can be supplemented in the future.

4.1. Summary of key findings

Descriptive statistics: The surveyed sample predominantly comprises participants between 25 and 34 years old (56%), male (71%) professionals with the majority in Developer (61%) roles. Most professionals are at the mid-level (41%) of their careers, hailing from medium-sized companies with 51-500 employees (22%), mainly in Russia (44%) and Italy (22%). The Business/IT sector (42%) had the highest representation.
Regarding the use of ChatGPT, it is clear that although its use is widespread, many use it discreetly, without the knowledge of their managers, or even in violation of company policy. Most benefit from the free version, citing possible barriers to the paid version or satisfaction with the features of the free offering.
General conclusions:

- ChatGPT's impact on Productivity: Our research reveals a positive impact of ChatGPT on productivity in software development. ChatGPT has showcased a moderate time-saving with a mean rating of 2.56 for more creative tasks and 2.25 for less creative tasks.
- Variations in ChatGPT's Influence on Software development stages: Findings of this study indicate that ChatGPT's influence on productivity varies across different stages of the software development process. Coding and Documentation stages report the highest perceived time-saving with means of 2.682 and 2.848 respectively. On the other hand, Deployment stands out as the stage with the least perceived benefit from ChatGPT, bearing a mean score of 1.756.

The second stage in the rating of those that benefit from ChatGPT less is Software Design. Its mean rating is 2.167. This could indicate that while ChatGPT provides value in many stages, the intuitive and visionary process of software design might be something where human expertise and creativity dominate.
Interesting to note that in stages Code Review and QA, the mode is 1 (No time saved), but their mean values are closer to 2.5. This indicates a polarization in user feedback: a significant number of users don't find any time-saving, while others might find considerable benefits.
However, a significant difference between the perceived influence depending on creativity of stages cannot be states, as, how it was already previously mentioned, means in two groups are close to the same value (2.25 and 2.56).

- Effectiveness for experienced professionals More experienced software professionals tended to benefit less from ChatGPT, reporting lower levels of productivity improvement. This suggests that experienced professionals are better equipped to proceed with their tasks without ChatGPT. However, they might anyway employ the tool to validate their decisions or to get alternative perspectives on challenging problems.

Correlations and respondent characteristics:

- Role & Perception:

Senior professionals see fewer benefits with ChatGPT, possibly relying more on experience and intuition.

As professionals climb the hierarchy scale, their evaluations of ChatGPT become more critical.

- Experience & Efficiency:

The more experienced the professional, the fewer the perceived advantage of ChatGPT, likely due to reliance on traditional methods.

- Company Size & Perception:

Employees in larger firms see fewer benefits from ChatGPT, perhaps due to established corporate workflows that overshadow new tools.

- Version & User Perception:

Users of newer or premium ChatGPT versions view it more positively, likely due to enhanced features.

Overall, while ChatGPT's utility is recognized across various users, its benefits appear to be influenced by the professional's career stage, experience, company size, and ChatGPT version. The strength of these correlations, while significant, remains mostly weak to moderate.

ANOVA Test Results:

- Overall Impact on Productivity:

No significant difference by Grade.

Significant difference by Age group, with older and younger groups perceiving higher productivity.

No significant difference by Company Type.

Significant difference between ChatGPT versions, with paid users reporting higher productivity.

- Speed of Completing Tasks with ChatGPT:

The difference across Grade is nearing statistical significance but isn't significant.

- Possibility to Perform Several Tasks at the Same Time:

The ability across different grades is nearing statistical significance but isn't significant.

- Feeling Less Tired with ChatGPT:

There's a statistically significant difference across grades with Intern/Junior employees feeling the least tired.

- Quality Improvement of Work with ChatGPT:

Statistically significant difference across grades. Younger or less experienced employees perceive a more positive impact.

- Errors Reduce with ChatGPT:

The difference in error reduction across grades is not statistically significant but is approaching significance.
Summary of results by hypothesis (Table 49).Table 49. Summary of results by hypothesis

Hypothesis	Results
H1.Less-experienced employees benefit more from using ChatGPT than those with higher levels of expertise.	+
H2. Employees having a lower grade benefit from ChatGPT more.	+
H3. There will not be a difference in influence on Productivity for employees work in large organizations or for employees of small firms / freelancers.	the difference is found, but it is unsignificant statistically
H4. ChatGPT's impact will vary for different software development stages.	+
H5. The impact of ChatGPT on software development productivity is more significant in less creative stages.	-
H6. Professional using paid version of ChatGPT (4.0) will record higher positive influence on their productivity.	+
H7. Younger participants will record more impact of ChatGPT as they tend to implement new technologies to life more actively.	+
H8. The impact of ChatGPT on software development productivity is higher in teams where the usage of ChatGPT is actively promoted by line manager (however, we supposed that the active promotion is rare).	could not be checked

Source: made by author

4.2. Discussion of results

The results of the study offer compelling evidence of the transformative impact of ChatGPT on software development productivity. The reduction in task completion time, combined with improved code quality, makes ChatGPT a good tool for software professionals seeking enhanced productivity.
Notably, the findings suggest that ChatGPT's strengths lie in the early stages of development, where generating ideas, documentation, and code play a pivotal role. By providing quick and contextually relevant suggestions, ChatGPT helps developers in formulating clear requirements and generating code, expediting the overall development process.
However, it is essential to recognize that ChatGPT's effectiveness may vary depending on the complexity of the task and the expertise. While experienced professionals might

leverage ChatGPT less efficiently, developers at all levels can still benefit significantly from its capabilities.
In the discussion of the results, it is important to acknowledge that direct comparisons with previous studies may not be feasible due to the uniqueness of this research. While there is a growing body of literature on AI in software development and productivity, the specific focus on ChatGPT's influence in this study sets it apart from existing research.
However, we can compare our results with MIT study of general productivity, not in software development, which results were released in July 2023 (Mashable, 2023). The experiment comprised 453 college-educated professional and randomly assigned half of the participants with ChatGPT after completing their first assignment. The assignments were writing-based tasks including press releases, short reports, and, emails, mimicking those that grant writers, marketers, consultants, data analysts, and HR professionals would do in their day-to-day work. The study found the group that was given access to ChatGPT decreased in time taken to accomplish a task by 11 minutes and increased in quality. This correlates with our findings.

4.3. Limitations of work

The research methodology and survey employed in this study have certain limitations that were discussed in 2.7 paragraph. Shortly, the sample size may not fully represent the entire software development industry, and self-reported data could introduce biases and inaccuracies. The cross-sectional design captures only a snapshot of experiences, and potential confounding factors were not controlled for. Additionally, subjectivity in data analysis may have influenced the findings to some extent. Despite these limitations, the study provides valuable insights into ChatGPT's impact on productivity in software development. Future research can explore alternative methods to further enhance our understanding of ChatGPT's role in this context.

4.4. Implications of the findings

The findings of our study have significant implications for both practitioners and researchers in the field of software development and artificial intelligence. These implications shed light on the potential benefits and challenges associated with the integration of ChatGPT into software development workflows.

1. Enhanced Productivity and Efficiency: The most prominent implication of our study is the undeniable positive impact of ChatGPT on software development productivity. ChatGPT offers professionals the opportunity to accomplish more in less time, leading to increased efficiency and quicker project delivery. By automating repetitive tasks and generating contextually relevant suggestions, ChatGPT empowers professionals to focus on higher-level problem-solving and creative aspects of software development.
2. Augmenting skills: The integration of ChatGPT in software development workflows has the potential to augment the skills of professionals across different experience levels. While experienced professionals can leverage ChatGPT to amplify their

capabilities and streamline their processes, less experienced developers can benefit from the guidance and support provided by the language model. This can lead to a more inclusive and collaborative work environment, where developers can learn from AI-driven insights and enhance their expertise.

3. Optimizing software Development stages: Our study highlights the potential of ChatGPT to optimize specific stages of software development, such as requirements gathering and code generation. By providing timely and accurate suggestions, ChatGPT can facilitate the formulation of clear and precise requirements, leading to improved project planning and reduced rework. Moreover, its code generation capabilities can accelerate the initial development phases, where speed and accuracy are paramount.
4. Adaptability across organizational sizes: The versatility of ChatGPT becomes apparent in its adaptability to different organizational sizes. Large enterprises can harness the power of ChatGPT to manage complex projects efficiently, while small businesses and freelancers can benefit from its time-saving capabilities to enhance their productivity. This adaptability makes ChatGPT a promising tool for organizations of all sizes, driving innovation and growth in the software development landscape.
5. Future research opportunities: The implications of our study open up numerous avenues for future research. Exploring the optimal ways to integrate ChatGPT with different development methodologies, investigating the long-term impact of AI-driven language models on software development practices, and addressing challenges related to data quality and bias in AI training data are all promising areas for further investigation.

4.5. Suggestions for future research

Looking ahead, several avenues for future research emerge from our study. Exploring the optimal integration of ChatGPT with different software development methodologies could provide insights into maximizing its benefits. Additionally, addressing concerns related to data privacy, security, and ethics in utilizing AI-driven language models is crucial for their responsible and ethical implementation.

Understanding the long-term impact of ChatGPT on software development practices and the evolving role of developers in this AI-augmented landscape presents a fascinating area for further investigation. Future research can also delve into the integration of ChatGPT with other AI technologies to create more powerful and synergistic AI-driven development environments.

Bibliography / References

1.1. Artificial Intelligence: evolution, capabilities and applications

1.1.1 Evolution of AI and its diverse applications

1. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
2. Gupta, S., Davidson, P., & Malick, R. (2018). The future of mobility: What's next? Deloitte Insights.
3. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 2(4), 230-243.
4. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
5. L., Rochester, N., & Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. AI Magazine, 27(4), 12-14.
6. Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., ... & Ueda, K. (2016). Cyber-physical systems in manufacturing. CIRP Annals, 65(2), 621-641.
7. Newell, A., & Simon, H. A. (1961). GPS, a program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), Computers and Thought (pp. 279-293). McGraw-Hill.
8. Arner, D. W., Barberis, J. N., & Buckley, R. P. (2016). The evolution of fintech: A new post-crisis paradigm? Georgetown Journal of International Law, 47(4), 1271-1319.
9. Gomez-Uribe, C. A., & Hunt, N. (2016, August 10). Artwork personalization at Netflix. Netflix Technology Blog. Retrieved from https://netflixtechblog.com/artwork-personalization-c589f074ad76
10. IBM Watson. (n.d.). Regulatory compliance. Retrieved from https://www.ibm.com/watson/industries/regulatory-compliance
11. Mitchell, T. M. (1997). Machine Learning. McGraw-Hill Education.
12. Siciliano, B., & Khatib, O. (Eds.). (2016). Springer Handbook of Robotics. Springer.
13. Jurafsky, D., & Martin, J. H. (2019). Speech and language processing (3rd ed.). Prentice Hall.

1.1.2. Introduction to AI language models and their capabilities

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
2. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need.
4. Jurafsky, D., & Martin, J. H. (2019). Speech and Language Processing (3rd ed.).

1.2. ChatGPT: A Powerful AI Language Model

1.2.1. Overview of ChatGPT and its capabilities

1. OpenAI. (2023). Introducing ChatGPT. Retrieved from https://openai.com/research/chatgpt

2. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2022). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.
3. OpenAI. (2023). OpenAI Codex. Retrieved from https://openai.com/blog/openai-codex/
4. Henceforth Solutions (2023). How does Chat GPT differ from other AI programs and what precisely does it do? Retrieved from https://medium.com/@henceforthsolution10/how-does-chat-gpt-differ-from-other-ai-programs-and-what-precisely-does-it-do-6802a0ffc9f3

1.2.2. Potential use cases of ChatGPT in various industries

1. Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. BenchCouncil transactions on benchmarks, standards and evaluations, 2(4), 100089.
2. Haman, M., & Školník, M. (2023). Exploring the capabilities of ChatGPT in academic research recommendation. Resuscitation.
3. Understanding ChatGPT capabilities. Retrieved from https://bootcamp.uxdesign.cc/understanding-chatgpt-capabilities-c382877200ee?gi=9c987ed21ddc

1.3. Productivity in Software Development

1.3.1. Defining productivity and its significance

1. Solow, R. M. (1957). Technical change and the aggregate production function. The Review of Economics and Statistics, 39(3), 312-320.
2. OECD. Defining and Measuring Productivity. Retrieved from https://www.oecd.org/sdd/productivity-stats/40526851.pdf
3. Syverson, C. (2011). What determines productivity? Journal of Economic Literature, 49(2), 326-365.
4. Porter, M. E. (1990). The competitive advantage of nations. Harvard Business Review, 68(2), 73-93.
5. Krugman, P. (1994). The Age of Diminished Expectations: U.S. Economic Policy in the 1990s. Cambridge, MA: MIT Press.

1.3.2. Metrics for measuring productivity

1. Syverson, C. (2011). What determines productivity? Journal of Economic Literature, 49(2), 326-365.
2. Bureau of Labor Statistics. (n.d.). Labor productivity and costs. Retrieved from https://www.bls.gov/lpc/
3. Baily, M. N., & Gordon, R. J. (1988). The productivity slowdown, measurement issues, and the explosion of computer power. Brookings Papers on Economic Activity, 1988(2), 347-431.
4. Solow, R. M. (1957). Technical change and the aggregate production function. The Review of Economics and Statistics, 39(3), 312-320.
5. Organisation for Economic Co-operation and Development. (2001). Measuring productivity - OECD manual: Measurement of aggregate and industry-level productivity growth. OECD Publishing.

6. Nishimizu, M., & Page, J. M. (1982). Total factor productivity growth, technological progress and technical efficiency change: Dimensions of productivity change in Yugoslavia, 1965-78. The Economic Journal, 92(368), 920-936.

1.3.3. Review of survey-based studies on AI and productivity

1. Damioli, G., Van Roy, V. & Vertesy, D. The impact of artificial intelligence on labor productivity. Eurasian Bus Rev 11, 1–25 (2021).
2. Czarnitzki, D., Fernández, G. P., & Rammer, C. (2022). Artificial Intelligence and Firm-Level Productivity. ZEW - Leibniz Centre for European Economic Research Discussion Paper No. 22-005. https://ssrn.com/abstract=4049824
3. Alderucci, D., Branstetter, L., Hovy, E., Runge, A., & Zolas, N. (2019). Quantifying the Impact of AI on Productivity and Labor Demand: Evidence from U.S. Census Microdata. Center for AI Analysis of Patents Carnegie Mellon University and NBER.

1.3.4. Review of empirical analysis of ChatGPT's impact on productivity

1. Noy, S., & Zhang, W. (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. Working Paper, MIT.

1.3.5. Measuring productivity in software development

1. Sudhakar, G. P. (2011). Techniques/Models for measuring Software Development Team Productivity. SJM, 7(1), 3-6. Retrieved from https://ssrn.com/abstract=2425392
2. Tomaszewski' P. (2006) Software Development Productivity Evaluation and Improvement for Large Industrial Projects.

1.4. AI's Impact on Productivity in Software Development

1.4.1. Understanding Software Development process

1. Klopper, R., Gruner, S., & Kourie, D. (2007), "Assessment of a framework to compare software development methodologies" Proceedings of the 2007
2. Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries,

1.4.2. The Software Development Process and ChatGPTs transformational influence

1. Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. Harvard Business Review, 96(1), 108-116.
2. Nair, S., Lyytinen, K., & Kannabiran, G. (2018). AI in requirements engineering. IEEE Software, 35(5), 69-75.
3. Raroque Ch, The Role of AI in Software Development: Trends, Statistics, and Growth. Retrieved from https://aloa.co/blog/the-role-of-ai-in-software-development-trends-statistics-and-growth
4. Padmanaban, P. H., & Sharma, Y. K. (2019). Implication of Artificial Intelligence in Software Development Life Cycle: A state of the art review. International Journal of Recent Research Aspects
5. Kumari, V., & Kulkarni, S. (2018). Use of Artificial Intelligence in Software Development Life Cycle Requirements and its Model. International Research Journal of Engineering and Technology

1. Sobania, D., Briesch, M., Hanna, C., & Petke, J. (2023). An analysis of the automatic bug fixing performance of ChatGPT.

2. Surameery, N. M. S., & Shakor, M. Y. (2023). Use chat gpt to solve programming bugs. International Journal of Information Technology & Computer Engineering (IJITC).
3. Johnson H., Educative. (2023). How ChatGPT can help devs' productivity. Retrieved from https://www.educative.io/blog/chatgpt-how-it-can-help-devs-productivity
4. Doglio F., Medium. (2023). How to use ChatGPT effectively to increase developer efficiency. Retrieved from https://webuters.medium.com/how-to-use-chatgpt-effectively-to-increase-developer-efficiency-f2d908832be1
5. Sacolick I., Infoworld. (2023). ChatGPT and software development. Retrieved from https://www.infoworld.com/article/3689172/chatgpt-and-software-development.html
6. Infinity Innovators, LinkedIn. (2023). How Chat GPT is changing the landscape of software development. Retrieved from https://www.linkedin.com/pulse/how-chat-gpt-changing-landscape-software-development/
7. Honeypot. (2023). How can ChatGPT help developers? Retrieved from https://cult.honeypot.io/reads/how-can-chatgpt-help-developers/
8. Chawla Sh., Coforge. (2023). 4 Ways ChatGPT can help accelerate the software development process. Retrieved from https://www.coforge.com/blog/4-ways-chatgpt-can-help-accelerate-the-software-development-process
9. Maxwell M., MakeUseOf. (2023). How ChatGPT can improve your programming skills. Retrieved from https://www.makeuseof.com/chatgpt-programming-practical-uses/

Chapter 2. Methodology and Approach

1. Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics. SAGE Publications Ltd.
2. Salkind, N. J. (2020). Statistics for People Who (Think They) Hate Statistics. SAGE Publications Inc.
3. Montgomery, D. C. (2017). Design and Analysis of Experiments. John Wiley & Sons Inc.
4. Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. Pearson.

Chapter 4. Discussion and Conclusion

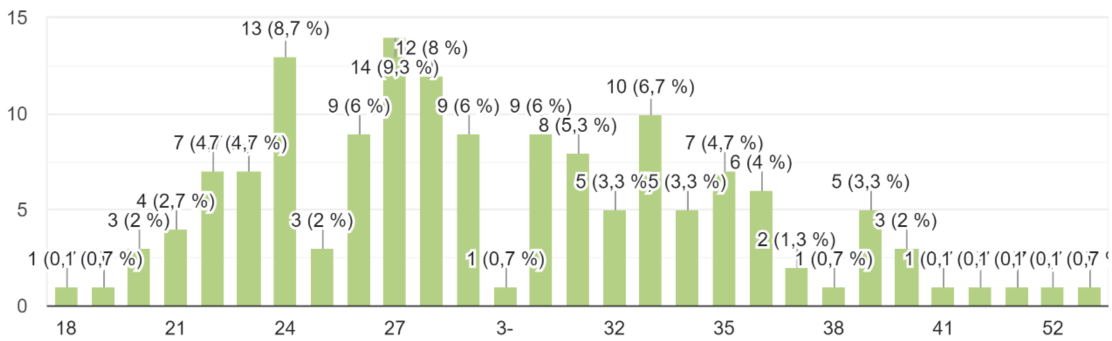
4.2. Discussion of results

1. MIT study: ChatGPT increases productivity for human workers (2023). Retrieved from https://mashable.com/article/mit-study-chatgpt-increases-productivity-decreases-inequality

Appendix 1. Survey design + results

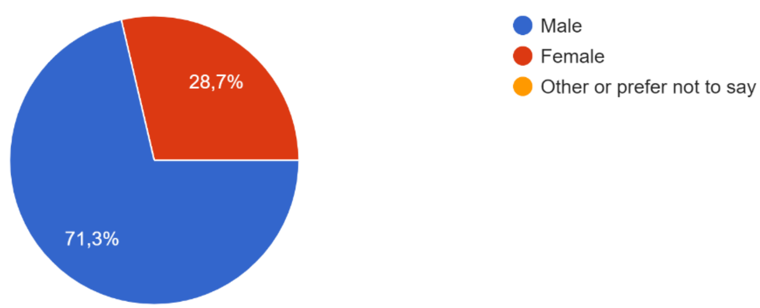
Age

150 ответов



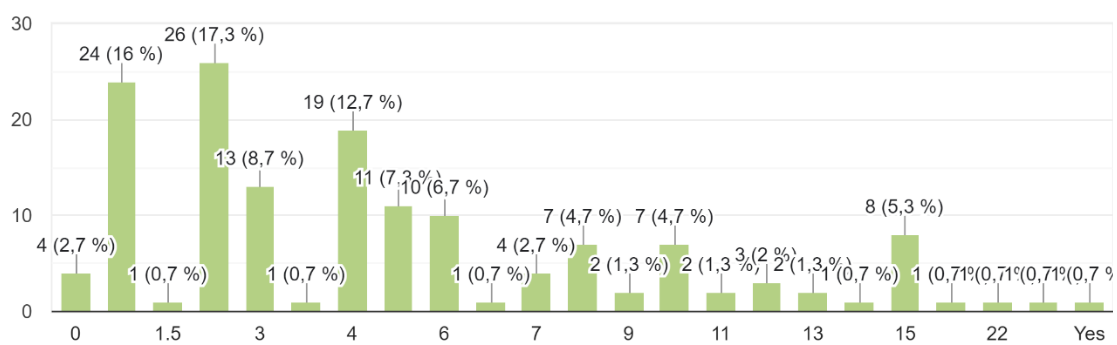
Gender

150 ответов



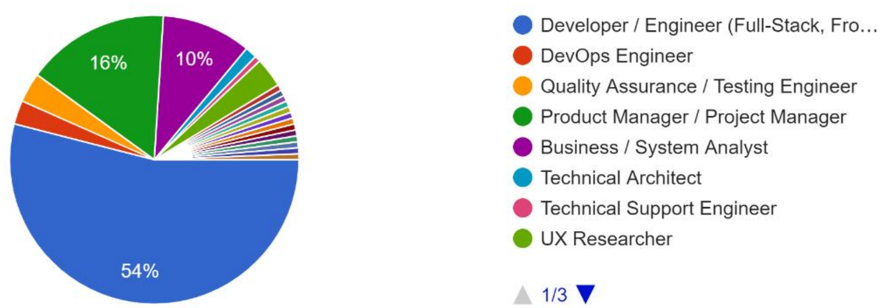
Years of experience in the software development industry

150 ответов



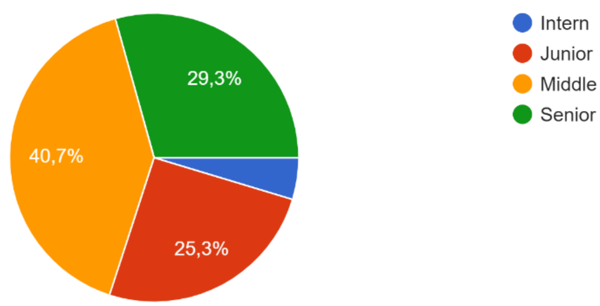
Current job role (if several, chose the main one)

150 ответов



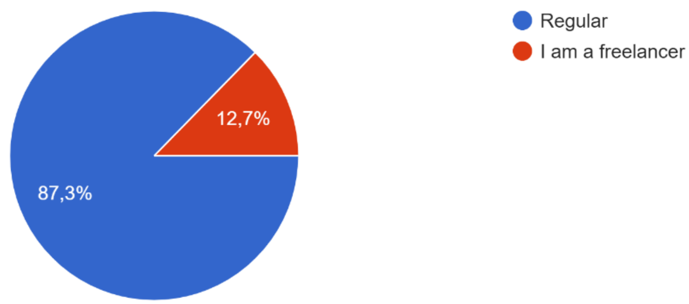
Current position level

150 ответов



Employment type

150 ответов



Company size (very roughly, employees)

150 ответов

0

50

10000

5000

500

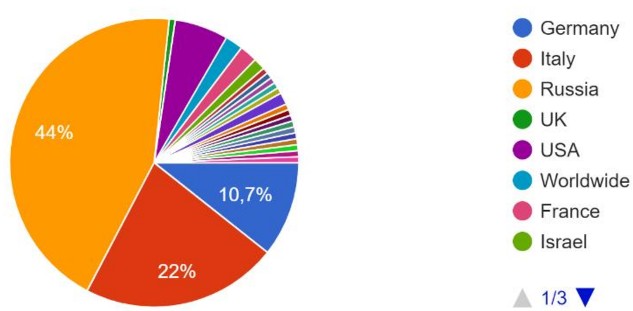
1000

200

20

Main country of company's operations

150 ответов

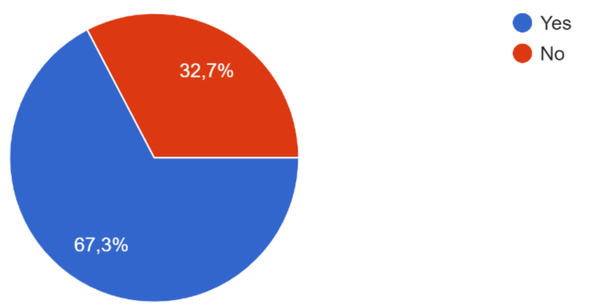


Main industry (of a company or of most of your clients, if you are a freelancer)

150 ответов

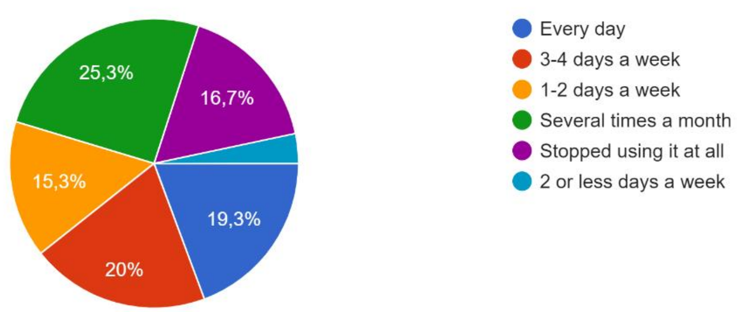
Are you using ChatGPT in your software development work regularly (at least once in two weeks)?

150 ответов

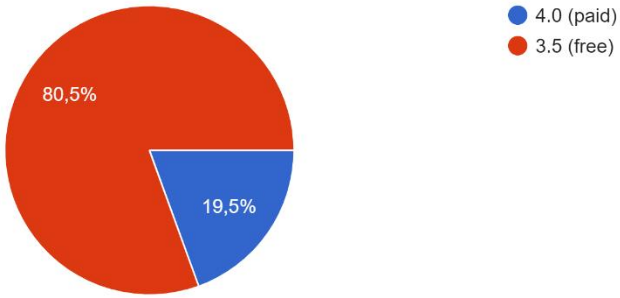


How frequently do you use ChatGPT for work? (considering, working week = 5 days)

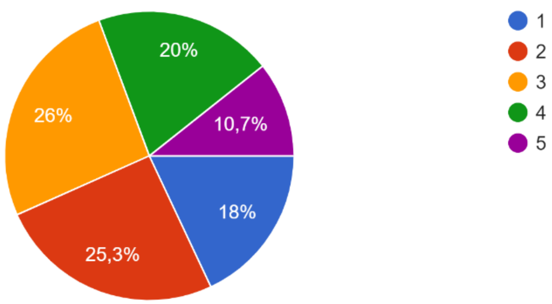
150 ответов



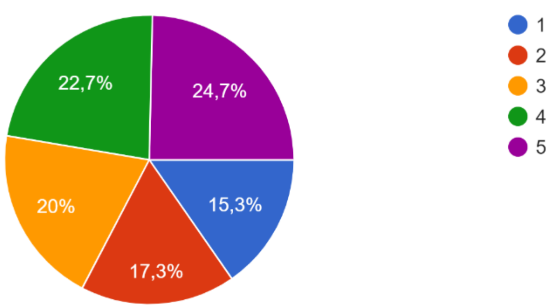
ChatGPT version you mainly use
149 ответов



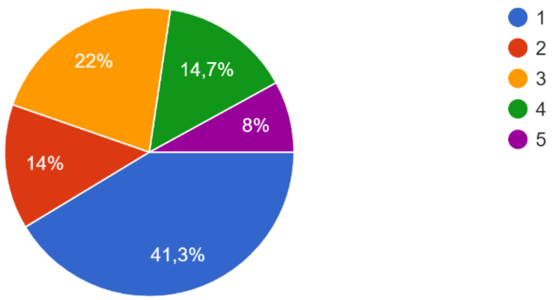
On a scale of 1 to 5, how do you rate the overall impact of ChatGPT on your productivity in software development? (1 = No impact, 5 = Extremely impactful)
150 ответов



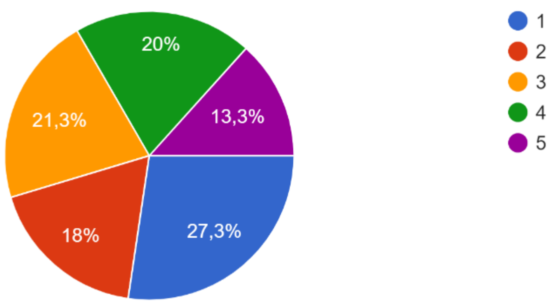
Rate the following statement on a scale of 1 to 5 (1 = Strongly disagree, 5 = Strongly agree): ChatGPT has helped me to complete tasks more quickly.
150 ответов



With the help of ChatGPT, I began to perform more tasks at the same time
150 ответов

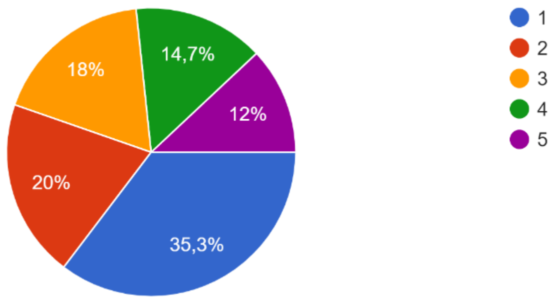


Using ChatGPT has allowed me to focus more on higher-level tasks and decision-making
150 ответов



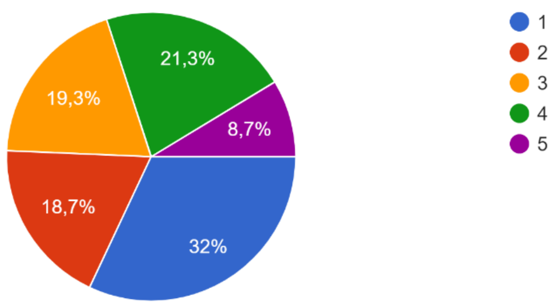
With the help of ChatGPT I began to feel less tired after completing tasks

150 ответов



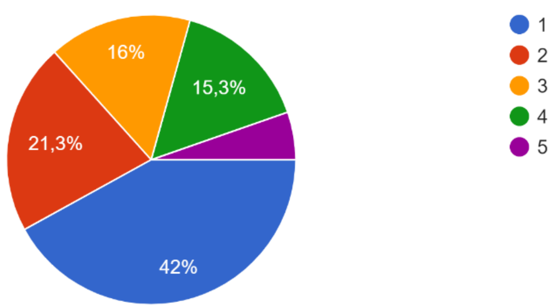
ChatGPT has improved the quality of my work

150 ответов



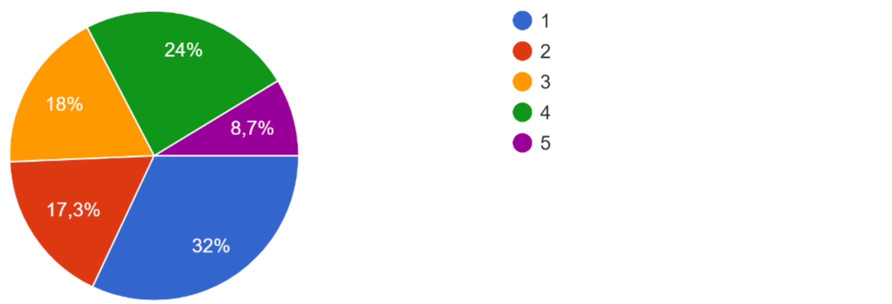
ChatGPT has reduced the number of errors in my work

150 ответов



ChatGPT has enhanced my problem-solving abilities

150 ответов



Please, for each stage estimate the percentage of time saved due to ChatGPT   Planning

150 ответов

Requirements analysis

150 ответов

Software design

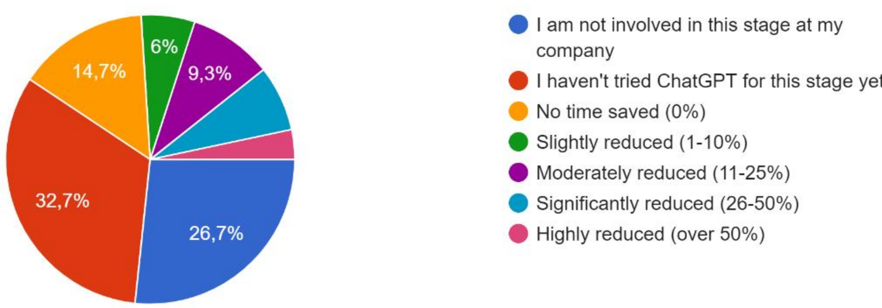
150 ответов

Coding

150 ответов

Code review

150 ответов



Testing and QA

150 ответов

Deployment

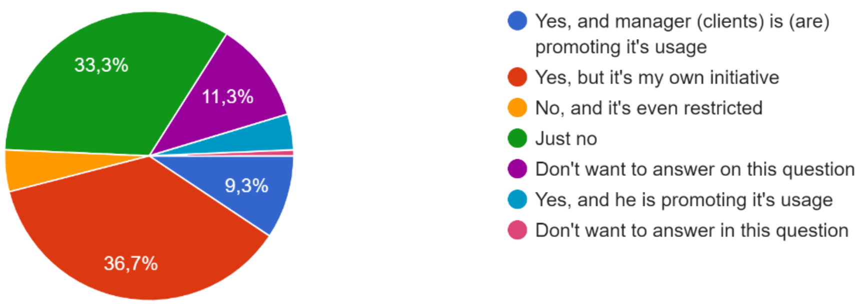
150 ответов

Maintenance and bug fixing

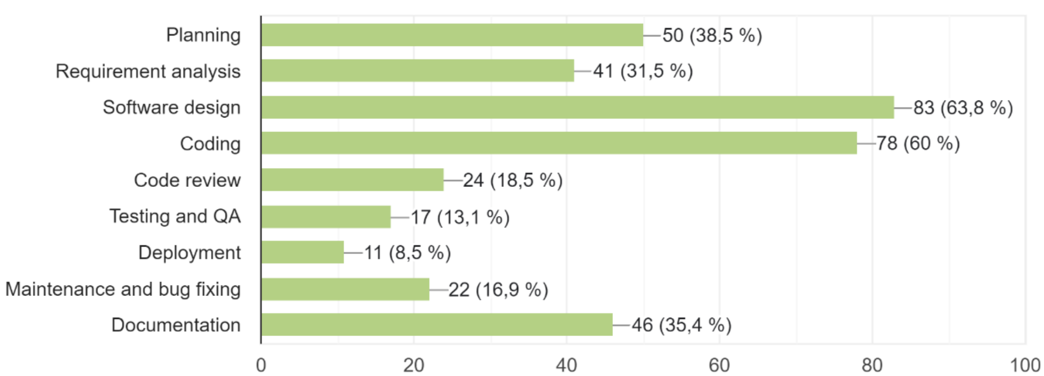
150 ответов

Documentation
150 ответов

Do(es) your line manager or your clients (if you are a freelancer) know(s) that you are using ChatGPT for work?
150 ответов

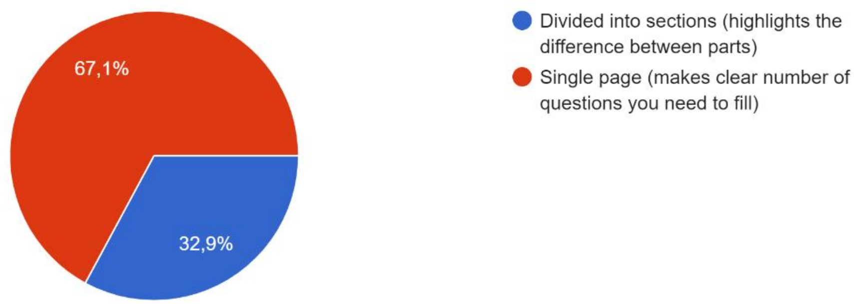


Name 4 or less stages that you consider the most 'creative'
130 ответов



*What type of a survey design would you prefer?

76 ответов



Space for your comments

22 ответа

I don’t have a lot of experience using chatgpt for coding, I used for other purposes, for example to explore vision and mission of other companies etc.
for programming purposes:
It can help you if you’re a freelancer that create mostly static websites and things like that, but for big projects that uses multiple languages and have a bigger complexity I think that is not so useful at the moment.

It's an amazing AI tool to do anything you can think of, but I think it's dangerous if we lose control of it, it's worth the life of the human species

For me, ChatGPT usually works as an improved Google. It doesn't automate my work, it just helps me to find answers faster

Ciao!

The level of understanding and the quality of responses it provides are truly impressive. ChatGPT has not only saved me time and effort but also expanded my knowledge and helped me explore new perspectives. I highly recommend giving it a try! It's an invaluable resource that can enhance your productivity and provide a delightful and insightful experience.

Appendix 2. Description of variables

Variable	Description	Comments
Age	Age	
Age_group	Age group	1. 18-24 2. 25-34 3. 35-44 4. 45 and above
Gender	Gender	
Experience	Years of experience in the software development industry	
Job_role	Current job role (if several, chose the main one)	
Grade	Current position level	1. Intern/Junior 2. Middle 3. Senior
Employment_type	Employment type	
Company_size	Company size (very roughly, employees)	
Company_type	Company type	0 freelance 1. Micro: 1-10 employees 2. Small: 11-50 employees 3. Medium: 51-500 employees 4. Large: 501-5,000 employees 5. Enterprise: 5,001-50,000 employees 6. Giant: Over 50,000 employees
Country	Main country of company's operations	
Industry	Main industry (of a company or of most of your clients, if you are a freelancer)	
Active_user	Are you using ChatGPT in your software development work regularly (at least once in two weeks)?	
Frequency	How frequently do you use ChatGPT for work?	

	(considering, working week = 5 days)	
ChatGPT_version	ChatGPT version you mainly use	
Overall_impact_productivity	On a scale of 1 to 5, how do you rate the overall impact of ChatGPT on your productivity in software development? (1 = No impact, 5 = Extremely impactful)	(1 = No impact, 5 = Extremely impactful)
Speed	Rate the following statement on a scale of 1 to 5 (1 = Strongly disagree, 5 = Strongly agree): ChatGPT has helped me to complete tasks more quickly.	1 = Strongly disagree, 5 = Strongly agree
More_tasks_same_time	With the help of ChatGPT, I began to perform more tasks at the same time	
Focus_higher_lvl	Using ChatGPT has allowed me to focus more on higher-level tasks and decision-making	
Less_tired	With the help of ChatGPT I began to feel less tired after completing tasks	
Quality_improvement	ChatGPT has improved the quality of my work	
Errors_reduce	ChatGPT has reduced the number of errors in my work	
Boost_problem_solving	ChatGPT has enhanced my problem-solving abilities	
Planning_1	Please, for each stage estimate the percentage of time saved due to ChatGPT Planning	1. No time saved (0%) - I haven't tried ChatGPT for this stage yet 2. Slightly reduced (1-10%) 3. Moderately reduced (11-25%) 4. Significantly reduced (26-50%) 5. Highly reduced (over 50%)
Requirements_1	Requirements analysis	
Software_design_1	Software design	
Coding_1	Coding	
Code_review_1	Code review	
Testing_QA_1	Testing and QA	
Deployment_1	Deployment	
Maintenance_bug_fix_1	Maintenance and bug fixing	
Documentation_1	Documentation	

Manager_knows	Do(es) your line manager or your clients (if you are a freelancer) know(s) that you are using ChatGPT for work?	

Appendix 3. Significance of data

Variable	Level	Counts	Total	Proportion	p
Planning_1	No time saved (0%)	20	69	0.29	< .001
	Slightly reduced (1-10%)	19	69	0.275	< .001
	Moderately reduced (11-25%)	15	69	0.217	< .001
	Significantly reduced (26-50%)	12	69	0.174	< .001
	Highly reduced (over 50%)	3	69	0.043	< .001
Requirements_1	No time saved (0%)	21	67	0.313	0.003
	Slightly reduced (1-10%)	24	67	0.358	0.027
	Moderately reduced (11-25%)	12	67	0.179	< .001
	Significantly reduced (26-50%)	9	67	0.134	< .001
	Highly reduced (over 50%)	1	67	0.015	< .001
Software_design_1	No time saved (0%)	19	66	0.288	< .001
	Slightly reduced (1-10%)	27	66	0.409	0.175
	Moderately reduced (11-25%)	12	66	0.182	< .001

	Significantly reduced (26-50%)	6	66	0.091	< .001
	Highly reduced (over 50%)	2	66	0.03	< .001
Coding_1	No time saved (0%)	17	109	0.156	< .001
	Slightly reduced (1-10%)	32	109	0.294	< .001
	Moderately reduced (11-25%)	34	109	0.312	< .001
	Significantly reduced (26-50%)	20	109	0.183	< .001
	Highly reduced (over 50%)	6	109	0.055	< .001
Code_review_1	No time saved (0%)	22	61	0.361	0.04
	Slightly reduced (1-10%)	9	61	0.148	< .001
	Moderately reduced (11-25%)	14	61	0.23	< .001
	Significantly reduced (26-50%)	11	61	0.18	< .001
	Highly reduced (over 50%)	5	61	0.082	< .001
Testing_QA_1	No time saved (0%)	16	50	0.32	0.015
	Slightly reduced (1-10%)	15	50	0.3	0.007

	Moderately reduced (11-25%)	10	50	0.2	< .001
	Significantly reduced (26-50%)	5	50	0.1	< .001
	Highly reduced (over 50%)	4	50	0.08	< .001
Deployment_1	No time saved (0%)	22	41	0.537	0.755
	Slightly reduced (1-10%)	11	41	0.268	0.004
	Moderately reduced (11-25%)	5	41	0.122	< .001
	Significantly reduced (26-50%)	2	41	0.049	< .001
	Highly reduced (over 50%)	1	41	0.024	< .001
Maintenance_bug_fix_1	No time saved (0%)	20	76	0.263	< .001
	Slightly reduced (1-10%)	27	76	0.355	0.015
	Moderately reduced (11-25%)	13	76	0.171	< .001
	Significantly reduced (26-50%)	13	76	0.171	< .001
	Highly reduced (over 50%)	3	76	0.039	< .001
Documentation_1	No time saved (0%)	15	79	0.19	< .001

Slightly reduced (1-10%)	16	79	0.203	< .001
Moderately reduced (11-25%)	24	79	0.304	< .001
Significantly reduced (26-50%)	14	79	0.177	< .001
Highly reduced (over 50%)	10	79	0.127	< .001

* Yellow is for values that are not significant

List of tables

List of figures

Declaration in lieu of oath

By Daria Glushkova

This is to confirm my Master Thesis was independently composed/ authored by myself, using solely the referred sources and support. I additionally assert that this thesis has not been part of another examination process

Глушкова

October 2023