Productivity vs. Purpose: Generative AI Enhances Task Performance but Reduces Meaningfulness in Programming

Maren Mehler
Technical University of Darmstadt, maren.mehler@tu-darmstadt.de

Kai Krautter
Harvard Business School, kaikrautter@g.harvard.edu

PERFORMANCE VS. PURPOSE: GENERATIVE AI ENHANCES TASK PERFORMANCE BUT REDUCES MEANINGFULNESS IN PROGRAMMING

Completed Research Paper

Maren F. Mehler, Technical University Darmstadt, Darmstadt, Germany, mehler@is.tu-darmstadt.de

Kai Krautter, Harvard Business School, Boston, United States, kaikrautter@g.harvard.edu

Abstract

Generative Artificial Intelligence (GenAI) has become widespread in daily work but present novel challenges for users as previously meaningful tasks can now be completed by GenAI. This study examines the impact of ChatGPT on task performance and perceived meaningfulness in two programming tasks. In an online experiment (n=161) assigning participants to coding or debugging tasks, with and without ChatGPT assistance, we found that using ChatGPT improved task performance, partially because the supported tasks are less difficult. However, using ChatGPT resulted in lower perceived meaningfulness, partly because participants considered the tasks less effortful. Notably, both tasks exhibited slightly different results, indicating that contextual factors may amplify or mitigate the effects. This study emphasizes the dual nature of GenAI integration, balancing enhanced performance with psychological impacts on users. Our findings offer insights for organizations and developers on integrating GenAI, highlighting the importance of incorporating efficiency gains with the meaningfulness of human work.

Keywords: Meaningfulness, Task Performance, Generative AI, ChatGPT, Online Experiment.

1 Introduction

In an era where technology has become an integral part of our daily lives, artificial intelligence (AI) facilitates various aspects of work (Brynjolfsson and Mitchell, 2017; Berente et al., 2021; Russell and Norvig, 2021). For example, AI is expected to create new roles with new tasks requiring skills such as creativity, problem-solving, or empathy (Bryant, 2023). As AI evolves, advances in generative AI (GenAI), especially in Large Language Models (LLMs) over the course of the last year, have resulted in tools such as ChatGPT, which are expanding the potential applications and fostering a new era of collaboration between AI and humans. In areas like customer service, where ChatGPT can provide responses, in marketing, where GenAI can be used to produce content such as images with Dall-E 3 (Northwest Executive Education, 2023; OpenAI, 2023) or GitHub Copilot in supporting coders in their daily work (GitHub Blog, 2023) new possibilities arise. So far, AI has been effective at performing routine tasks (e.g., translating (DeepL, 2023), freeing up human capacity for more engaging, creative and meaningful tasks. However, the introduction of content creation tools—GenAI—using LLMs such as ChatGPT and image generation tools like Dall-E 3 is changing the landscape and raising intriguing questions about the extent to which AI can now handle tasks that were once considered the exclusive domain of human creativity (Dwivedi et al., 2023).

Thus, the main objective of this paper is to examine whether individuals (i) improve their performance in complex activities with GenAI assistance (hereafter referred to as *task performance*) and (ii) how the usage of (generative) AI impacts how they feel about their work (hereafter referred to as *perceived* *meaningfulness*). Specifically, we are interested in how employees perceive their work as meaningful, given that finding passion and meaningfulness in their work is a top priority for many employees (Cech, 2021; Jachimowicz and Weisman, 2022). Academic research has supported the notion that finding one's work meaningful can increase motivation and, consequently, job performance (Allan, 2017; Jachimowicz et al., 2018). To better understand the interaction between AI and human employees, it is thus crucial not only to consider how people perform at work but also how they feel about their work.

We test our research questions in the context of programming—an area within corporate organizations that stands to derive significant advantages from the assistance of GenAI. In addition, a fusion of technical expertise and creative problem-solving is paramount in this domain. Specifically, ChatGPT can aid in various aspects, such as code generation, error detection, and testing (Perkel, 2023). While many possibilities to adopt ChatGPT and ease work exist (Dwivedi et al., 2023), ChatGPT cannot serve as a substitute for the primary contributions of programmers. Still, it can effectively streamline and enhance their work processes (Larson, 2023). Therefore, we examine whether AI assistance can increase task performance in programming tasks and whether it affects how people feel about their work (i.e., how meaningful they perceive their work). In an online experiment we asked 161 participants to either complete coding or debugging tasks with or without ChatGPT's assistance. We observed that partly due to less difficulty higher task performance could be attained with the assistance of ChatGPT. Conversely, the use of ChatGPT resulted in partially viewing the tasks as less effortful and thus lower perceived meaningfulness was found among participants. However, as both tasks results are slightly different, various contextual factors may influence our observed effects. Consequently, our results suggest that ChatGPT's assistance influences individuals' perceptions of their work, a factor that should be taken into consideration in future decisions regarding human-AI collaboration.

The paper is structured as follows: In the next chapter, we outline the background and address the development of AI into GenAI. We also focus on how task performance and perceived meaningfulness have been measured and used in prior research. In the third chapter, we derive our hypotheses for our online experiments. The fourth chapter presents our methodology, where we describe the design of the experiment and the data collection process. This is followed by the analysis of the results in the fifth chapter, and a discussion concludes this paper with the findings, our contributions, limitations, and future research in the last chapter.

2 Theoretical Background and Hypotheses Development

This chapter is divided into two parts. First, we provide a definition and overview of AI, focusing on the characteristics of GenAI and ChatGPT, as well as the influence of GenAI on programming. Then, we will present an overview of the measurements of work engagement: task performance, i.e., how people perform at work, and perceived meaningfulness, i.e., how they feel about their work. Additionally, we point out previous work on both aspects in combination with ChatGPT.

2.1 Artificial intelligence

In the field of AI, the emergence of GenAI, exemplified by systems like ChatGPT, has introduced in a paradigm shift in the perception and interaction with AI technologies. Previously, AI and its subcategory, machine learning (ML), were learning algorithms making decisions or predictions based on recognized patterns in the data (Mitchell, 1997; Brynjolfsson and Mitchell, 2017; Russell and Norvig, 2021). However, with GenAI and LLMs, tools such as ChatGPT, Bing Chat, or, specifically for programming tasks, GitHub Copilot, it became possible to generate new data from identified patterns rather than solely analyzing the data (Dwivedi et al., 2023; Teubner et al., 2023). Tasks that were once assumed to be exclusively within the realm of human capability were already being replaced before the advent of GenAI (Bankins and Formosa, 2020), and this trend may further intensify now.

Prior research on human-AI collaboration has primarily operated under the assumption that AI can make decisions but not generate content (e.g., Fügener et al., 2021; Jussupow et al., 2021). It has been demonstrated that when the AI is correct, it can enhance human performance (Fügener et al., 2021;

Boyacı, Canyakmaz, and de Véricourt, 2023), emphasizing the potential benefits of humans and AI working together on tasks. However, this research also indicates that humans tend to prefer challenging tasks and are hesitant to delegate them to AI systems, even though this delegation could improve performance (Fügener et al., 2022). These initial findings suggest that how humans feel about their work is often more important to them than how they perform.

With regard to programming, ChatGPT or more specific tools such as GitHub Copilot offer programmers the means to simplify their daily work with high-quality output (GitHub Copilot, 2023; Perkel, 2023). Programmers can now use AI to draft code, debug it, run tests, or receive explanations (E. Chen et al., 2023; Liu et al., 2023; Surameery and Shakor, 2023). As a result, these new possibilities are reshaping the way programmers work, which, in turn, impacts their performance and how they perceive the meaningfulness of their work. Consequently, ChatGPT has the potential to fundamentally transform how humans interact with technology (Haleem, Javaid, and Singh, 2022).

2.2 Engaging in work

Engagement in daily work is twofold. On the one hand, it needs to be measured by employers which is mainly done through performance measurements. On the other hand the work is assessed by how employees feel towards it – finding meaning in their work.

2.2.1 Performance

Task performance can be defined as the effectiveness with which a person carries out an activity, often pertaining to a task related to their job (Borman and Motowidlo, 1993, 1997). These activities have a direct connection to organizational outcomes, as they either transform resources into products and services or support such processes (Motowidlo and Scotter, 1994). Performance can generally encompass various aspects, such as the quality, quantity, and relevance of service (Makki and Abid, 2017). The specific interpretation of performance may vary based on the particular field or area of interest. In our context, we define task performance based on the quality of execution of a programming assignment. This can be subjectively evaluated by the person performing the assignment (Goodman and Svyantek, 1999), but it can also be assessed more objectively by analyzing the individual task output (see Benlian, 2015), which in our case refers to the final code snippets. Since measuring performance is very complex (Ramos-Villagrasa et al., 2019), we decided to use both.

In AI research, performance measurements, such as accuracy, are frequently used to demonstrate the effectiveness of algorithms (Russell and Norvig, 2021), similar to research in human-AI collaboration, where performance is measured to indicate its improvement when AI is employed (e.g., Fügener et al., 2021; Boyacı, Canyakmaz, and de Véricourt, 2023). Noy and Zhang (2023) have presented a working paper demonstrating that productivity can be enhanced when ChatGPT is used. They also find evidence that ChatGPT can boost job satisfaction. Gilardi, Alizadeh, and Kubli (2023) went one step further and demonstrated that ChatGPT outperforms crowd workers by an average of 25 percentage points for annotation tasks. Nonetheless, the measurement of performance using GenAI is still in its early stages. For instance, Chen et al. (2023) conducted an empirical study for evaluating text quality, and they identified asking ChatGPT for an explicid score as the most reliable and effective one.

Assessing how well a person performs in a programming assignment can be done by evaluating code quality. Various metrics, such as complexity, can be used for this purpose (Baggen et al., 2012). More specifically, there are many metrics falling into several categories. For example, the code's degree of complexity can be measured through code comprehensibility, while the absence of errors can be assessed through software bug prediction (Nuñez-Varela et al., 2017). Additionally, coding standards and code reviews can enhance quality (Li and Prasad, 2005; Boogerd and Moonen, 2008; Stegeman, Barendsen, and Smetsers, 2014; Kononenko, Baysal, and Godfrey, 2016). GenAI opens up new possibilities in supporting these efforts. For example, it has been demonstrated that ChatGPT can match other models and is even superior to standard program repair techniques in debugging (Surameery and Shakor, 2023).

Thus, the following hypotheses address the research question of whether AI affects task performance. It is not a novelty that others, for example via outsourcing (e.g., Elmuti, Grunewald and Abebe, 2010), or nowadays AI can support humans and take over work. Research has shown, for example, that when AI is correct, it can enhance human performance (Fügener et al., 2021; Boyacı, Canyakmaz, and de Véricourt, 2023). In this context, AI can also handle very complex tasks, such as assisting with programming (e.g., E. Chen et al., 2023; Liu et al., 2023; Surameery and Shakor, 2023). Therefore, we assume that the code quality of a programming task will improve if GenAI is used to support it. This can be assessed in two ways. Firstly, through changes in the interaction with technology (Haleem, Javaid, and Singh, 2022), we assume that participants will perceive an increase in their own task performance (self-reported). Secondly, we aim to demonstrate this by measuring code quality using more objective metrics. Based on previous research on the use of ChatGPT in programming (e.g., Perkel, 2023) and in the evaluation of texts (Chen et al., 2023), we expect that ChatGPT itself should be able to evaluate the task performance from participants' code snippets (AI-rated). Therefore, we hypothesize:

H1: Task performance increases in tasks with assistance of generative AI.

2.2.2 Meaningfulness

In addition to task performance, it is important to consider how people feel about their work concerning a programming use case. Nowadays, people aim to find meaning in their daily work (Cech, 2021; Jachimowicz and Weisman, 2022). Meaningfulness, or meaningful work, is defined as work that is worthwhile to oneself or personally significant (Allan, 2017). The benefits of meaningful work tasks can result in more motivation and, thus, higher performance of individuals, from which organizations can profit (Allan, 2017).

However, within the context of Information Systems (IS) research, the concept of meaningful work has often been relegated to describing results or emphasizing the significance of research findings. A structured literature review in the field of IS reveals a gap in the consideration of meaningfulness as a distinct construct. Despite the plethora of research in IS on well-being, only a limited number of studies have delved into the conceptualization and measurement of meaningfulness. Barkhi and Kao (2011) explore the psychological dimension of meaningfulness in the context of decision-making within Group Decision Support Systems (DSS). They find that users with higher levels of psychological meaningfulness make better decisions when they clearly understand the goal. Ke and Zhang's (2011) investigation into Open-source Software projects demonstrates that meaningfulness has a slightly negative influence on performance. The work of Kisekka and Goel (2022) introduces the concept of job meaningfulness as a factor influencing job performance, particularly during extreme events. Finally, Liu et al. (2017) see meaningful engagement in the gamification context as consisting of two elements, experiential and instrumental outcomes, thus also not directly measuring meaningfulness.

Therefore, it is essential to note that the term "meaningfulness" has not been consistently defined and utilized across various studies within the IS domain. The heterogeneity in the conceptualization of perceived meaningfulness underscores the need for a comprehensive and unified understanding of this construct to facilitate more coherent and comparable research outcomes.

Nonetheless, there have been ongoing discussions across disciplines about whether AI can increase or decrease perceived meaningfulness. AI, in general, can either replace human work, create new forms of collaboration, or amplify workers' skills (Bankins and Formosa, 2023). If AI automates human labor, the question arises whether humans lose the meaning in their work (Lysova et al., 2023) or if AI can relieve humans from repetitive or even dull tasks, thus increasing perceived meaningfulness (Bankins and Formosa, 2023). Especially with regard to the issue that people cannot always maintain the same passion for their work every day (Bredehorst et al., 2023), GenAI might either support individuals by reducing cognitive load or decrease perceived meaningfulness because the work feels less meaningful for them (Gielnik et al., 2015). As a result, the question remains about how the interaction changes when GenAI performs tasks that previously triggered creativity and added meaning to the job (Bankins and Formosa, 2023). We will aim to examine these questions in the current research.

Thus, the following hypotheses address the research question of whether AI affects the perceived meaningfulness of work concerning a programming use case. As concluded by Noy and Zhang (2023), the use of GenAI can impact job satisfaction. On the other hand, GenAI will automate human work, thus, meaning in their work might be lost (Lysova et al., 2023). In addition, programming is a rather creative task and not a repetitive one that can be automated to relieve humans (Bankins and Formosa, 2023). Although existing research does not provide a clear direction, we anticipate that the use of GenAI will generally have a negative effect on the perception of meaningfulness in programming. Therefore, we hypothesize:

H2: Perceived meaningfulness decreases in tasks with assistance of generative AI.

Finally, the next two hypotheses pertain to the mediation of our direct effects through perceived effort and task difficulty. It should logically follow that tasks that are easier to complete would also result in higher task performance. However, the concept of perceived meaningfulness adds complexity to the scenario. As technology can take over mundane task, it should feel more meaningful for individuals as they need to exert less effort to complete the same task (Bailey et al., 2019; Bankins and Formosa, 2023). However, research in psychology has indicated that the exertion of increased effort can sometimes make tasks feel more meaningful for individuals (Gielnik et al., 2015; Inzlicht and Campbell, 2022; Campbell, Chung and Inzlicht, 2023; Mortimer, 2023). Additionally, research on human-AI collaboration suggests that humans prefer challenging tasks and do not delegate them to AI systems (Fügener et al., 2022). Thus, if AI can relieve humans from repetitive or even monotonous tasks (Bankins and Formosa, 2023), lower task difficulty and reduced effort invested may actually increase performance but decrease perceived meaningfulness as challenging tasks are reduced. Therefore, we hypothesize:

H3: Assistance of generative AI predicts higher performance by reducing task difficulty.

H4: Assistance of generative AI predicts lower perceived meaningfulness by reducing perceived effort.

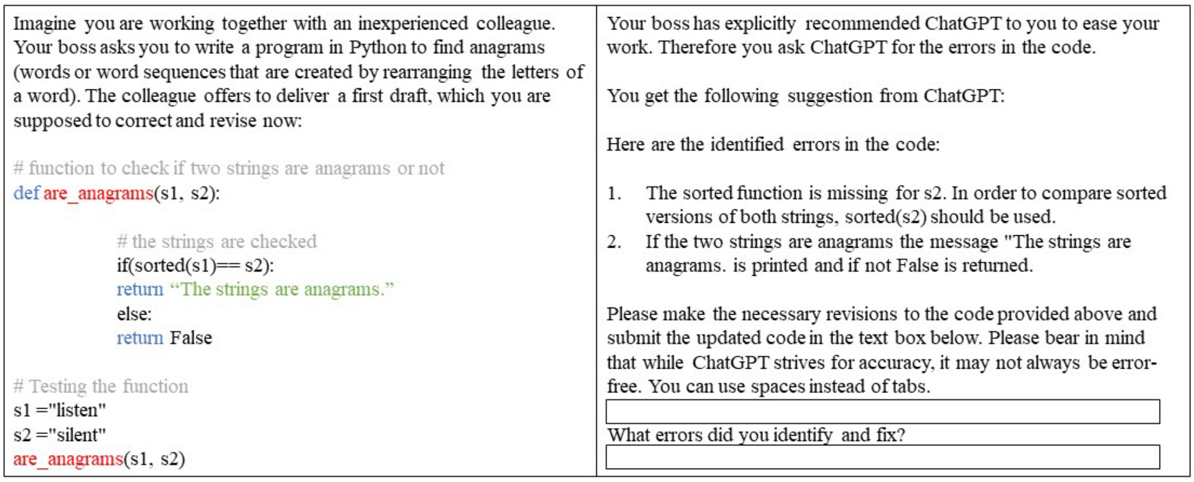
3 Methodology

To obtain our data, we conducted an online experiment on Prolific Academic involving experienced Python programmers. Our primary goal was to examine (i) the performance of programmers and (ii) the perceived meaningfulness of their work when instructed to complete tasks with the assistance of AI. Online experiments are well-suited for manipulating specific conditions, enabling us to evaluate our measures of interest after each task. This approach is commonly employed in comparable contexts within the field of IS (e.g., Turel, Yuan and Connelly, 2008).

3.1 Research design and measurements

To assess the effect of GenAI on task performance and perceived meaningfulness, we instructed participants to undertake two basic programming tasks in Python, one with AI assistance and one without AI assistance (randomly assigned). We chose ChatGPT as the GenAI that assists our participants, as it is well-known by them. We hypothesized that engaging in everyday work tasks with AI, such as programming, would alleviate task difficulty, thereby enhancing task performance (Fügener et al., 2021; Boyacı, Canyakmaz, and de Véricourt, 2023). However, it also reduces the required effort, decreasing participants' perceived meaningfulness (Gielnik et al., 2015).

For higher levels of external validity, we created two groups with different types of typical programming questions and randomly assigned participants to one of them. They were either asked to write code from scratch (coding group) or instructed to find mistakes in existing code (debugging group). The tasks within each group were typical for learning programming, such as writing code for Fibonacci sequences or finding anagrams (see Figure 1). All tasks were developed in collaboration with a computer science professor and an experienced computer science student. Afterward, we piloted the tasks and the survey with five Ph.D. students in IS with substantial programming experience. Based on their feedback, we adjusted the wording of the tasks and survey questions to enhance the study's comprehensibility.


Imagine you are working together with an inexperienced colleague. Your boss asks you to write a program in Python to find anagrams (words or word sequences that are created by rearranging the letters of a word). The colleague offers to deliver a first draft, which you are supposed to correct and revise now:

```
# function to check if two strings are anagrams or not
def are_anagrams(s1, s2):

        # the strings are checked
        if(sorted(s1)== s2):
        return "The strings are anagrams."
        else:
        return False

# Testing the function
s1 ="listen"
s2 ="silent"
are_anagrams(s1, s2)
```

Your boss has explicitly recommended ChatGPT to you to ease your work. Therefore you ask ChatGPT for the errors in the code.

You get the following suggestion from ChatGPT:

Here are the identified errors in the code:

1. The sorted function is missing for s2. In order to compare sorted versions of both strings, sorted(s2) should be used.
2. If the two strings are anagrams the message "The strings are anagrams. is printed and if not False is returned.

Please make the necessary revisions to the code provided above and submit the updated code in the text box below. Please bear in mind that while ChatGPT strives for accuracy, it may not always be error-free. You can use spaces instead of tabs.

What errors did you identify and fix?

Figure 1. Example task in debugging group with ChatGPT.

The experiment was structured as follows (see Figure 2): After participants agreed to a privacy policy, each participant was randomly assigned to one of the two groups: either coding from scratch or finding mistakes. Participants were asked to submit their code (completely new code for the coding group or improved code for the debugging group). If applicable, participants had to provide their prompts and the output of ChatGPT. Following each task, participants answered questions concerning their task performance (adapted from Goodman and Svyantek, 1999), perceived meaningfulness (adapted from May, Gilson and Harter, 2004), perceived effort (adapted from Paas, 1992), perceived difficulty (on a 5-point Likert scale; adapted from Ribeiro and Yarnal, 2010), and the extent to which they utilized ChatGPT or other tools (e.g., Google or similar search engines) in the task (as a manipulation check). These measures were adapted from existing surveys to ensure validity and reliability. Finally, after completing the two tasks, participants answered demographic questions regarding their age and gender (see Table 1), as well as control variables related to their attitude toward AI, experience with coding, and coding enjoyment (see the Appendix for the constructs and items). Most variables were assessed on a 7-point Likert scale ranging from 1 = *Completely Disagree* to 7 = *Completely Agree*. All variables were integrated into an online survey tool and were marked as mandatory.

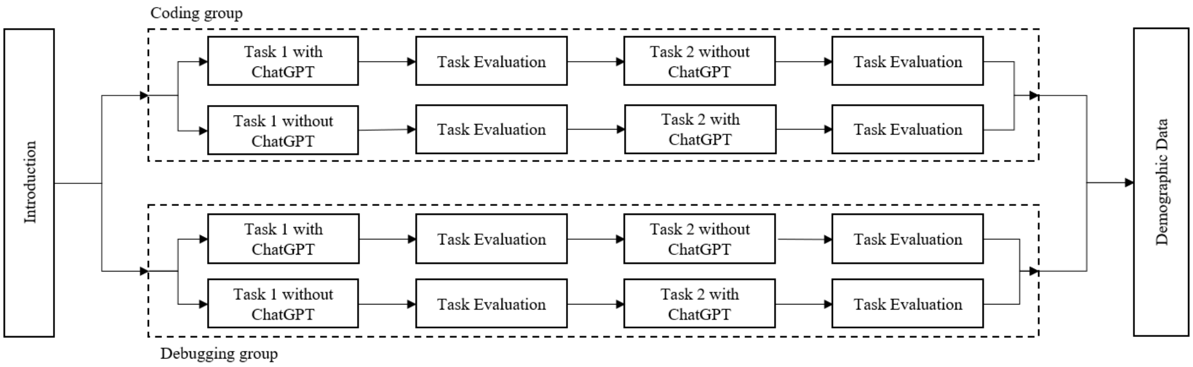


Figure 2. Experiment structure (double arrow indicates random allocation).

In addition to assessing task performance based on the subjective responses from participants, we also evaluated task performance by reviewing the code and identifying errors in the task of error detection. Consequently, similar to Chen et al. (2023) for text quality, we requested ChatGPT to rate the quality of the code on a scale from 1 to 10, considering various aspects of code quality, such as whether the code fulfills the task, its efficiency, adherence to coding standards, and so on. After ChatGPT rated all 322 coding snippets, the authors randomly reviewed the code and ratings and generally concurred with the assigned scores.

3.2 Data collection

After developing and testing our experiment, we administered it and collected data via Prolific Academic, allowing us to filter for our target group, which consisted of individuals fluent in English and experienced software developers in Python. We successfully collected the intended sample size of 160 participants. We excluded and replaced three participants who did not have access to ChatGPT, which resulted in a final sample size of 161. The participants were rewarded with about 10 €/h and completed the experiment in a mean time of 16 minutes. Most participants fell within the range of 18-34, and the gender distribution was balanced, as we had specified during the recruitment process on Prolific (see Table 1). Additionally, participants in our sample were relatively experienced coders (*M* = 5.26, *SD* = 1.12) who also enjoyed coding (*M* = 5.58, *SD* = 1.16). Furthermore, participants had favorable attitudes toward AI (*M* = 5.57, *SD* = 1.18).

Demography	**Categories**	**Frequency**	**Demography**	**Categories**	**Frequency**
Age	18-24	61 (38%)	**Gender**	Female	79 (49%)
	25-34	72 (45%)		Male	79 (49%)
	35-44	18 (11%)		Non-Binary	3 (2%)
	45-54	8 (5%)			
	<55	2 (1%)			

Table 1. Demographic data (age and gender) of the participants.

4 Results

In the following section, we will first present the results for the coding group, followed by the results for the debugging group. For both groups, we calculate Cronbach's alpha, composite reliability (CR), and average variance extracted (AVE) to confirm the internal consistency of our measures (Fornell and Larcker, 1981). Afterward, we will conduct paired *t*-tests to examine whether participants performed better and reported higher levels of perceived meaningfulness in the tasks with the assistance of ChatGPT. Finally, we will perform a mediation analysis to assess whether the relationship between the manipulation (task completion with and without ChatGPT's assistance) and task performance, as well as perceived meaningfulness, is mediated by perceived task difficulty and the effort invested in the task.

4.1 Coding group

4.1.1 Reliability and validity

Table 2 depicts the factor loadings for each variable, which are higher than the factor loadings on all other constructs, indicating convergent and discriminant validity. In addition, discriminant validity is verified, as the square root of AVEs (bold on the diagonal) is greater than all inter-construct correlations (Fornell and Larcker, 1981). All constructs' values for Cronbach's alpha, CR, and AVE were above the thresholds of 0.7, 0.7, and 0.5, respectively, thus confirming internal consistency (Hair Jr. et al., 2006, 2017). For this group, no significant correlation between task performance and perceived meaningfulness was found.

Construct	Factor loadings	Cronbach's Alpha	CR	AVE	1	2
1. Self-reported Performance	0.835-0.901	0.888	0.920	0.743	**0.862**	
2. Perceived Meaningfulness	0.834-0.925	0.964	0.959	0.800	-0.026	**0.893**

*Table 2. Reliability, validity, and correlation matrix (bold numbers are the square root of AVEs; significant correlations (p < .05) are marked with *).*

4.1.2 Hypotheses test

We conducted paired *t*-tests between our manipulation (task completion with and without ChatGPT's assistance) and self-reported task performance and AI-rated task performance (H1), as well as perceived meaningfulness (H2). In line with H1, our results indicate that the use of ChatGPT predicted a higher self-reported task performance for participants (M = 6.25, SD = 0.87) compared to self-reported task performance without the assistance of ChatGPT (M = 6.09, SD = 1.01), $t(77) = 1.8$, $p < .050$. To measure the effect size, we also calculate Cohen's d. Here, the effect size d = 0.20 indicates a small effect. Also, our results indicate that the use of ChatGPT predicted a higher AI-rated task performance for participants (M = 7.47, SD = 0.66) compared to AI-rated task performance without the assistance of ChatGPT (M = 6.47, SD = 2.14), $t(77) = 3.9$, $p < .001$ with an effect size of d = 0.44, indicating a medium effect. Thus, H1 was supported by our data. In line with H2, our results indicate that the use of ChatGPT predicted a lower perceived meaningfulness for participants (M = 3.88, SD = 1.58) compared to perceived meaningfulness without the assistance of ChatGPT (M = 4.35, SD = 1.49), $t(77) = -3.7$, $p < .001$ with an effect size d = 0.42, indicating a medium effect. Thus, H2 was also supported by our data.

Next, we conducted a structural model analysis based on our hypotheses H3 and H4 using a bootstrapping procedure involving 5,000 subsamples in SmartPLS 4 (Kushary, Davison and Hinkley, 2000) (see Figure 3 for the complete model). In particular, we estimated a mediation model from our manipulation (the assistance of ChatGPT or not) to our outcome variables, including self-reported and AI-rated task performance, as well as perceived meaningfulness, through the two mediators perceived task difficulty and effort. Our model demonstrates that the independent variables and the mediation account for a significant portion of the variance in the dependent variables (see R^2 for each construct). We assessed all hypotheses by examining the path coefficient and its significance using p-values.

First, we found a significant relationship between the use of ChatGPT and task difficulty (b = -0.630, $p < .001$), indicating that using ChatGPT resulted in decreased task difficulty for the participants. Similarly, they perceived their tasks as less effortful when they used ChatGPT (b = -0.786, $p < .001$).

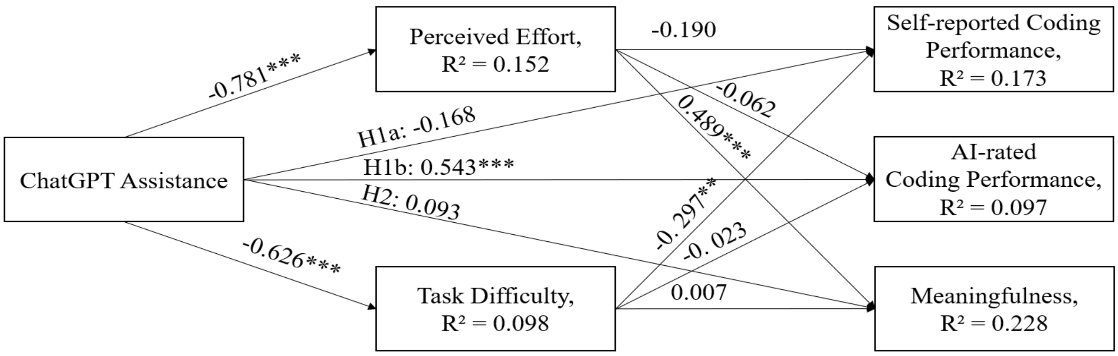


Figure 3. Results of the structural model for the coding group (p < .05, ** p < .01, *** p < .001).*

Second, task difficulty was negatively related to self-reported task performance (b = -0.297, $p < .010$) but showed no significant connection with AI-rated task performance (b = -0.023, $p > .050$). Task difficulty was also not significantly related to perceived meaningfulness (b = 0.007, $p > .050$). On the other hand, perceived effort did not have a significant effect on self-reported task performance (b = -

0.190, $p > .050$) or AI-rated task performance ($b = -0.062$, $p > .050$) but had a positive effect on perceived meaningfulness ($b = 0.489$, $p < .001$). As a result, the indirect effect between the use of ChatGPT and outcome variables through task difficulty was only significant for self-reported task performance ($b = 0.186$, $p < .050$). In contrast, the indirect effect between the use of ChatGPT and outcome variables through perceived effort was only significant for perceived meaningfulness ($b = -0.382$, $p < .001$). Thus, H3 was partially supported and H4 was supported by our data. Finally, when including the mediating variables, the direct effect between the use of ChatGPT and self-reported task performance became non-significant ($b = -0.168$, $p > .050$). Likewise, the direct effect between the use of ChatGPT and perceived meaningfulness was also non-significant ($b = 0.093$, $p > .050$). However, the direct effect between the use of ChatGPT and AI-rated task performance remained significant ($b = 0.543$, $p < .001$), as we did not identify a significant mediator for this relationship. Taken together, the effect of the use of ChatGPT on self-reported task performance was fully mediated via perceived task difficulty, and the effect of the use of ChatGPT on perceived meaningfulness was fully mediated via perceived effort.

4.2 Debugging group

4.2.1 Reliability and validity

Again, Table 3 shows the factor loadings for each variable, which are higher than the factor loadings on all other constructs, indicating convergent and discriminant validity. In addition, discriminant validity is verified as the square root of AVEs (bold on the diagonal) are greater than all inter-construct correlations (Fornell and Larcker, 1981). All constructs' values for Cronbach's alpha, CR, and AVE were above the thresholds of 0.7, 0.7, and 0.5, respectively, thus confirming internal consistency (Hair Jr. et al., 2006, 2017). Here, a positive and significant correlation between task performance and perceived meaningfulness can be found.

Construct	**Factor loadings**	**Cronbach's Alpha**	**CR**	**AVE**	1	2
1. Self-reported Performance	0.828-0.915	0.910	0.930	0.769	**0.877**	
2. Perceived Meaningfulness	0.864-0.916	0.965	0.976	0.785	0.202*	**0.886**

*Table 3. Reliability, validity, and correlation matrix (bold numbers are the square root of AVEs; significant correlations (p < .05) are marked with *).*

4.2.2 Hypotheses test

In line with our procedure for the first group, we conducted paired *t*-tests between our manipulation (task completion with and without ChatGPT's assistance), self-reported task performance and AI-rated task performance (H1), and perceived meaningfulness (H2). In contrast to the first task, our results indicate that the assistance of ChatGPT is not significant regarding a higher self-reported task performance for the participants ($M = 5.99$, $SD = 0.92$) compared to self-reported task performance without the assistance of ChatGPT ($M = 5.98$, $SD = 1.03$), $t(82) = 0.1$, $p > .050$ with an effect size of d = 0.01, indicating a very small effect. However, our results indicate that the use of ChatGPT predicted a higher AI-rated task performance for participants ($M = 5.20$, $SD = 2.72$) compared to AI-rated task performance without the assistance of ChatGPT ($M = 3.77$, $SD = 1.93$), $t(82) = 3.6$, $p < .001$ with an effect size of d = 0.40, indicating a medium effect. Thus, H1 is only partially supported by our data. In contrast to the first group, H2 was not supported by our data, given that the assistance of ChatGPT did not indicate a higher perceived meaningfulness for the debugging group ($M = 4.69$, $SD = 1.66$) compared to perceived meaningfulness without the assistance of ChatGPT ($M = 4.75$, $SD = 1.59$), $t(82) = -0.7$, $p > .050$ with an effect size of d = 0.08, indicating a very small effect.

Again, we conducted a structural model analysis based on our hypotheses H3 and H4 (see Figure 4 for the complete model). Our model demonstrates that the independent variables and the mediation account for a portion of the variance in the dependent variables (see R^2 for each construct). We assessed all hypotheses by examining the path coefficient and its significance using p-values.

First, we found a significant relationship between the use of ChatGPT and task difficulty (b = -0.293, $p < .050$) again, indicating that using ChatGPT resulted in decreased task difficulty for the participants. Similarly, participants perceived their tasks as less effortful when they used ChatGPT (b = -0.301, $p < .050$).

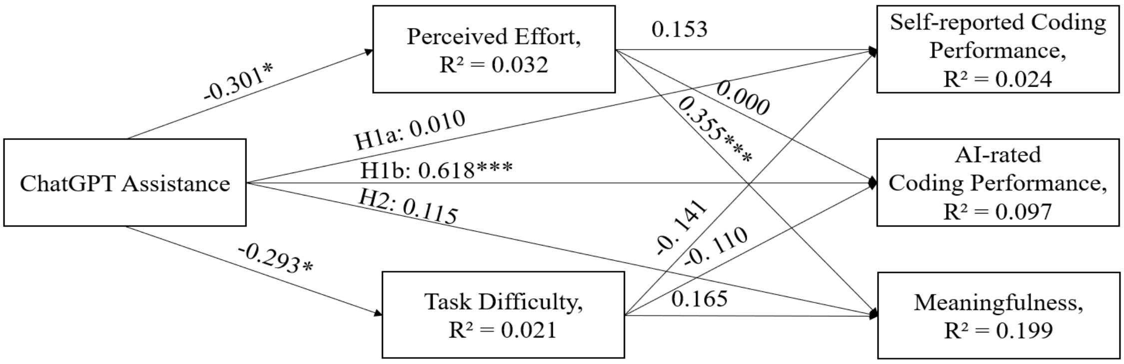


Figure 4. Results of the structural model for the debugging group (p < .05, ** p < .01, *** p < .001).*

Second, the task difficulty had no significant effect on either the self-reported task performance (b = -0.141, $p > .050$), AI-rated task performance (b = -0.110, $p > .050$), or perceived meaningfulness (b = 0.165, $p > .050$). Thus, the indirect effect of ChatGPT assistance via task difficulty to self-reported task performance is not significant (b = 0.041, $p > .050$). Also, the perceived effort did not have a significant effect on self-reported task performance (b = 0.153, $p > .050$) or AI-rated task performance (b = 0.000, $p > .050$) but had a positive significant effect on perceived meaningfulness (b = 0.355, $p < .001$). However, the indirect effect of ChatGPT assistance via perceived effort to perceived meaningfulness was not significant (b = -0.107, p = .099). Since the effect between ChatGPT assistance and perceived effort and perceived effort and perceived meaningfulness are quite small, the indirect effects between the use of ChatGPT and the outcome variables through the mediators were non-significant. Thus, H3 was not supported and H4 was partially supported by our data. Finally, when including the mediating variables, the direct effect between the use of ChatGPT and self-reported task performance was still non-significant (b = 0.010, $p > .050$). Likewise, the direct effect between the use of ChatGPT and perceived meaningfulness was also not significant (b = 0.115, $p > .050$). However, the direct effect between the use of ChatGPT and AI-rated task performance remained significant (b = 0.618, $p < .001$). Taken together, our results for the second task were descriptively similar to those in the first task. However, we did not observe as many significant effects, suggesting that contextual factors between the two different tasks may either amplify or mitigate the effects we observed.

In summary, we can support H1 with the coding task and partially with the debugging task. We can accept H2 for the coding, but not the debugging task. H3 is partially supported by the coding, but not the debugging task and finally, we can accept H4 for coding and partially for the debugging task.

5 Discussion

AI has become an integral part of everyday life (Brynjolfsson and Mitchell, 2017; Russell and Norvig, 2021; Berente et al., 2021), especially with the emergence of GenAI tools such as ChatGPT (Dwivedi et al., 2023; Teubner et al., 2023). While the enhancement of decision-making and task performance through human-AI collaboration is well-documented (e.g., Fügener et al., 2021; Boyacı, Canyakmaz and de Véricourt, 2023), the psychological well-being and perceptions of its users also can have an influence on the adoption of such technologies. Therefore, it is important to understand what influences the use of GenAI beyond task performance. Surprisingly, limited attention has been directed towards psychological constructs such as perceived meaningfulness in the context of AI. Initial studies, like those by Noy and Zhang (2023), suggest a positive correlation between ChatGPT use and job satisfaction, which we intend to delve deeper into. After all, higher perceived meaningfulness can lead to greater motivation and, therefore, higher organizational performance (Allan, 2017). In line with this, our primary goal of this experiment was to examine the evaluation of coders' task performance and their perception of meaningfulness when engaged in tasks with AI support. Such insights can pave the way for more holistic strategies in AI adoption, ensuring that technological advancements are harmoniously aligned with human well-being and purpose. To address our research question and test the subsequent hypotheses, we divided our participants into two groups with two different assignments. This allows us to test our research questions with two different scenarios, reflecting ChatGPT's multifaceted programming potential. Thus, the coding group had to solve coding tasks, and the debugging group had to perform error detection tasks. Both groups had to fulfill one task with and the other without ChatGPT assistance, which served as our primary manipulation in this experiment.

The first part of our research question centered around the task performance augmentation with AI assistance. Task performance was measured in two ways: self-reported by the participants of the experiment and AI-rated with ChatGPT. Consistent with previous studies on human-AI collaboration (Fügener et al., 2021; Boyacı, Canyakmaz and de Véricourt, 2023) we found that participants performed higher in their tasks when they had assistance from ChatGPT. This was shown in both groups through significant *t*-tests between the tasks solved with ChatGPT and without its assistance. In addition, in our structural model, a direct effect was particularly strong for the more objective AI-rated task performance, but we also found an indirect effect via task difficulty for self-reported task performance in the coding group. Furthermore, participants reported reduced effort and task complexity when assisted by ChatGPT. These observations underscore the potential of GenAI in simplifying tasks. Thus, we conclude that the assistance of ChatGPT improves the performance in complex activities such as coding.

Regarding the second part of our research question on how AI assistance impacts the sense of perceived meaningfulness in people's work, we identified significant differences with *t*-tests only for the use of ChatGPT in the coding group, indicating that participants reported lower levels of perceived meaningfulness when they used ChatGPT; the *t*-test in the debugging group, on the other hand, was not significant. However, the mediating effect through perceived effort emerged as significant for both groups, suggesting that for employees to rate a task as meaningful, they also need to feel that they have invested effort in it. This is intriguing because task performance is only mediated by task difficulty. Importantly, our findings are in line with emerging research in psychology that has indicated that increased effort can sometimes make tasks feel more meaningful for individuals (Gielnik et al., 2015; Inzlicht and Campbell, 2022; Campbell, Chung, and Inzlicht, 2023; Mortimer, 2023). Thus, our findings suggest that when people perceive tasks with ChatGPT assistance as less effortful, the meaningfulness decreases accordingly.

It is noteworthy that we also found nuanced differences between the two programming tasks. While not conclusive, the directionality of the relationship between perceived effort and self-reported task performance varied between the groups. For instance, a negative relationship was found in the coding group, whereas in the debugging group, a positive one was observed. Additionally, the influence of task difficulty on self-reported task performance in the second group is no longer significant. This may be attributed to the perception that the task is easier, as the ChatGPT support has a slightly less pronounced impact on the task difficulty and perceived errors. Consequently, the support of the GenAI no longer makes as significant a difference for the participants. Alternatively, it is plausible that coding is perceived as more creative, and thus, the use of GenAI has a greater influence than in debugging, in particular error-finding, which might be more repetitive and mundane.

5.1 Contributions

With our experiment, we can offer various theoretical contributions. As technology exceeds its traditional function of automating routine tasks and enters areas of creativity, it is crucial to assess the relationship between AI and human work. First, we align with the existing literature on human-AI collaboration and show that the use of AI results in higher task performance (e.g., Fügener et al., 2021; Boyacı, Canyakmaz, and de Véricourt, 2023). This is especially relevant for the specific coding tasks that have been investigated to a lesser extent in the past, as previous literature has mainly focused on AI for decision-making. We can, therefore, extent existing human-AI research with showing that this is also relevant for GenAI in coding tasks thus generalize on the one hand that GenAI leads to higher task performance and, on the other hand, that AI has a positive influence on these coding tasks. Second, our study aims to bridge the gap between the IS literature and meaningful work. Our paper addresses this void by systematically examining the role of perceived meaningfulness, particularly in the context of programming tasks. Thus, we also assess the psychological outcomes of how GenAI changes the way we work combining IS research with known phenomena from psychology (see Bailey et al., 2019). This is especially relevant as Gen Z, for example, rate meaningfulness as one of the main factors to take a new job (McKinsey, 2022). Depending on the task, we observed a decrease in perceived meaningfulness with ChatGPT assistance. It is possible that AI reduces the effort we have to invest in our work, leading to a perception of our work as less meaningful. Since these results are mixed, we can only provide an initial assessment of the influence of GenAI on perceived meaningfulness. Third, in the past, AI has been less capable of handling highly creative tasks (Dwivedi et al., 2023). This is now changing with GenAI. It was previously assumed that creative tasks, in particular, lead to a higher level of perceived meaningfulness, which is also demonstrated here to some extent. Thus, our research indicates that the focus should be on replacing or collaboratively solving relatively easy, boring, or repetitive tasks with (generative) AI, while leaving complex or creative tasks to humans that allow them to find meaning in their work (Cech, 2021; Jachimowicz & Weisman, 2022). Perhaps even new areas of work that are complex and particularly suitable for humans to take responsibility in the future will develop.

In addition to theoretical contributions, we also provide several practical ones. Firstly, companies should ensure that employees' perceived meaningfulness remains intact when working with GenAI. After all, perceived meaningfulness leads to more motivation, which, in turn, results in higher company performance (Allan, 2017). To achieve meaningfulness, it is important to consider which tasks can be performed collaboratively with AI, such as replacing routine tasks while jointly working on complex tasks or letting humans perform creative tasks. Additionally, the literature on job crafting offers models that can provide valuable guidance. Secondly, companies should allow and promote the use of GenAI within the company due to the increase in task performance caused by human-AI interaction. This is because higher individual performance contributes to higher company performance. Finally, while we recommend that employees be encouraged to use AI, companies should consider and evaluate which tools are most beneficial. ChatGPT can be utilized for texts or coding, but more specialized AI systems, such as GitHub Copilot for coding or DALL-E 3 for generating images, exist. In this way, GenAI can potentially reduce cognitive load and, therefore, create benefits for employees using such technologies. However, questions about the privacy and security of data entered into GenAI tools remain, as well as concerns about extending existing models with the company's own data.

5.2 Limitations and future research

Like any research, our study is not without limitations. First, our sample size is quite young, although software engineers are in general younger this is a limitation where future research might look into the effect on the effect on older people. While we utilized two different programming assignments in our experiment, they were still very specific. Thus, future research should aim to replicate our findings with additional (programming) tasks and different GenAI tools. Tools like Microsoft 365 Copilot might be interesting as it can change the entire scope of work, even the administrative area. In addition, we suggest exploring contextual effects that may either enhance or diminish the strength of our observed effects. Furthermore, although ChatGPT's responses were artificially predetermined in the debugging group, enhancing comparability might not accurately represent the interaction and, therefore, could impact the perceived meaningfulness. However, in the coding group, participants were required to interact with ChatGPT in real time during the coding task. The interaction, however, could only be assessed through control questions on the prompts and the output of ChatGPT. Finally, a high level of perceived meaningfulness might be attributed to the novelty of the situation. Although we made sure that the participants had experience with ChatGPT the novelty may diminish influencing the perceived meaningfulness – similar to experiences on outsourcing – if interactions with GenAI become routine. For instance, the literature on passion demonstrates significant fluctuations throughout daily work (Bredehorst et al., 2023). Therefore, future studies could potentially benefit from a later time of measurement and should be duly conducted.

Appendix

Table A shows the constructs, items, and control variables measured during the experiment:

Construct	**Item Description**	**Source**
Task Performance (self-reported)	I achieved the objectives of the task. I met criteria for performance. I fulfilled all the requirements of the task. I performed well in the overall task by carrying out tasks as expected.	Goodman and Svyantek (1999)
Perceived Meaningfulness	This task was very important to me. Completing this task was personally meaningful to me. The task I just did was worthwhile. This task was significant to me. The work I did on this task was meaningful to me. I feel that the work I did on this task was valuable.	May, Gilson and Harter (2004)
Perceived Effort	Please choose the level of effort that applied to you during this task.	Paas (1992)
Task Difficulty	How did you perceive the difficulty of this task?	Ribeiro and Yarnal (2010)
Attitude toward AI	How would you assess your attitude towards ChatGPT (or similar systems)?	-
Experience with Coding	How would you rate your coding experience (e.g., writing code for similar tasks like in our study)?	-
Coding Enjoyment	How enjoyable do you find coding (e.g., writing code for similar tasks like in our study)?	-

Table A. Constructs, control variables, and their items measured in the experiment.

Acknowledgement

This research and development project is/was funded by the German Federal Ministry of Education and Research (BMBF) within the “Innovations for Tomorrow’s Production, Services, and Work” Program (funding number 02L19C150) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

References

Allan, B.A. (2017) 'Task significance and meaningful work: A longitudinal study', *Journal of Vocational Behavior*, 102, pp. 174–182.

Apple (2023) *Siri*. Available at: https://www.apple.com/de/siri/ (Accessed: 10 October 2023).

Baggen, R. et al. (2012) 'Standardized code quality benchmarking for improving software maintainability', *Software Quality Journal*, 20(2), pp. 287–307.

Bailey, C., et al. (2019) 'The five paradoxes of meaningful work: Introduction to the special issue 'meaningful work: Prospects for the 21st century'', *Journal of Management Studies*, 56(3), pp. 481-499.

Bankins, S. and Formosa, P. (2020) 'When AI meets PC: exploring the implications of workplace social robots and a human-robot psychological contract', *European Journal of Work and Organizational Psychology*, 29(2), pp. 215–229.

Bankins, S. and Formosa, P. (2023) 'The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work', *Journal of Business Ethics* [Preprint].

Barkhi, R. and Kao, Y.-C. (2011) 'Psychological climate and decision-making performance in a GDSS context', *Information & Management*, 48(4–5), pp. 125–134.

Benlian, A. (2015) 'IT Feature Use over Time and its Impact on Individual Task Performance', *Journal of the Association for Information Systems*, 16(3), pp. 144–173.

Berente, N. et al. (2021) 'Managing Artificial Intelligence', *MIS Quarterly*, 45(3), pp. 1433–1450.

Boogerd, C. and Moonen, L. (2008) 'Assessing the value of coding standards: An empirical study', in *2008 IEEE International Conference on Software Maintenance*, Beijing, China, pp. 277–286.

Borman, W.C. and Motowidlo, S.J. (1993) 'Expanding the criterion domain to include elements of contextual performance', in *Personnel selection in organizations*.

Borman, W.C. and Motowidlo, S.J. (1997) 'Task Performance and Contextual Performance: The Meaning for Personnel Selection Research', *Human Performance*, 10(2), pp. 99–109.

Boyacı, T., Canyakmaz, C. and de Véricourt, F. (2023) 'Human and Machine: The Impact of Machine Input on Decision Making Under Cognitive Limitations', *Management Science* [Preprint].

Bredehorst, J. et al. (2023) 'The Challenge of Maintaining Passion for Work over Time: A Daily Perspective on Passion and Emotional Exhaustion', *Organization Science*, pp. 1-23.

Bryant, K. (2023) *How AI Will Impact the Next Generation Workforce*. Available at: https://www.forbes.com/sites/kalinabryant/2023/05/31/how-ai-will-impact-the-next-generation-workforce/ (Accessed: 13 October 2023).

Brynjolfsson, E. and Mitchell, T. (2017) 'What Can Machine Learning Do? Workforce Implications', *Science*, 358(6370), pp. 1530–1534.

Campbell, A. V., Chung, J. M. H., and Inzlicht, M. (2022) 'Meaningfulness of effort: Deriving purpose from really trying'. *PsyArXiv*.

Cech, E. (2021) *The trouble with passion: How searching for fulfillment at work fosters inequality*, University of California Press.

Chen, E. et al. (2023) 'GPTutor: A ChatGPT-Powered Programming Tool for Code Explanation', in N. Wang et al. (eds) *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky. Cham: Springer Nature Switzerland (Communications in Computer and Information Science)*, pp. 321–327.

Chen, Y. et al. (2023) 'Exploring the use of large language models for reference-free text quality evaluation: A preliminary empirical study', *arXiv preprint* arXiv:2304.00723 [Preprint].

DeepL (2023) *Home*. Available at: https://www.deepl.com/de/translator (Accessed: 10 October 2023).

Diener, E. et al. (2010) 'New well-being measures: Short scales to assess flourishing and positive and negative feelings', *Social indicators research*, 97, pp. 143–156.

Dwivedi, Yogesh K. et al. (2023) 'Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy', *International Journal of Information Management*, 71.

Elmuti, D., Grunewald, J., & Abebe, D. (2010). Consequences of outsourcing strategies on employee quality of work life, attitudes, and performance. *Journal of business strategies*, 27(2), 177-203.

Fornell, C., and Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), pp. 39–50.

Fügener, A. et al. (2021) 'Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI', *MIS Quarterly*, 45(3), pp. 1527–1556.

Fügener, A. et al. (2022) 'Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation', *Information Systems Research*, 33(2), pp. 678–696.

Gielnik, M.M. et al. (2015) '"I put in effort, therefore I am passionate": Investigating the path from effort to passion in entrepreneurship', *Academy of Management Journal*, 58(4), pp. 1012–1031.

Gilardi, F., Alizadeh, M. and Kubli, M. (2023) 'ChatGPT outperforms crowd-workers for text-annotation tasks', *arXiv preprint* arXiv:2303.15056 [Preprint].

GitHub Blog (2023) *Research: Quantifying GitHub Copilot's impact on code quality.* Available at: https://github.blog/2023-10-10-research-quantifying-github-copilots-impact-on-code-quality/ (Accessed: 07 November 2023).

GitHub Copilot (2023) *Your AI pair programmer*. Available at: https://github.com/features/copilot (Accessed: 13 October 2023).

Goodman, S. A. and Svyantek, D. J. (1999) 'Person–organization fit and contextual performance: Do shared values matter', *Journal of Vocational Behavior*, 55(2), 254-275.

Hair Jr., J.F. et al. (2006) *Multivariate Data Analysis*. 6th edn. Pearson Prentice Hall.

Hair Jr., J.F. et al. (2017) *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 2nd edn. SAGE Publications Inc.

Haleem, A., Javaid, M. and 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.

Howe, L.C., Jachimowicz, J.M. and Menges, J.I. (2021) *Your Job Doesn't Have to Be Your Passion*. Available at: https://hbr.org/2021/06/your-job-doesnt-have-to-be-your-passion (Accessed: 30 October 2023).

Inzlicht, M., and Campbell, A. V. (2022), 'Effort feels meaningful.' *Trends in Cognitive Sciences*.

Jachimowicz, J. M. et al. (2018) 'Why grit requires perseverance and passion to positively predict performance', *Proceedings of the National Academy of Sciences*, 115(40), pp. 9980-9985.

Jachimowicz, J. M., and Weisman, H. (2022) 'Divergence between employer and employee understandings of passion: Theory and implications for future research', *Research in Organizational Behavior*, 100167.

Jussupow, E. et al. (2021) 'Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence', *Information Systems Research*, 32(3), pp. 713–735.

Ke, W. and Zhang, P. (2011) 'Effects of Empowerment on Performance in Open-Source Software Projects', *IEEE Transactions on Engineering Management*, 58(2), pp. 334–346.

Kim, S.H. and Kwon, J. (2019) 'How Do EHRs and a Meaningful Use Initiative Affect Breaches of Patient Information?', *Information Systems Research*, 30(4), pp. 1184–1202.

Kisekka, V. and Goel, S. (2022) 'An Investigation of the Factors that Influence Job Performance During Extreme Events: The Role of Information Security Policies', *Information Systems Frontiers* [Preprint].

Kononenko, O., Baysal, O. and Godfrey, M.W. (2016) 'Code review quality: how developers see it', in *Proceedings of the 38th International Conference on Software Engineering*. ICSE'16, Austin, Texas, pp. 1028–1038.

Kushary, D., Davison, A. C., and Hinkley, D. V. (2000) 'Bootstrap Methods and Their Application', *Technometrics*, 42(2), 216.

Larson, Z. (2023) *ChatGPT Isn't Coming for Your Coding Job*. Available at: https://www.wired.com/story/chatgpt-coding-software-crisis/ (Accessed: 10 October 2023).

Li, X. and Prasad, C. (2005) 'Effectively teaching coding standards in programming', in *Proceedings of the 6th Conference on Information Technology Education*. SIGITE05, Newark, USA, pp. 239–244.

Liu et al. (2017) 'Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems', *MIS Quarterly*, 41(4), pp. 1011–1034.

Liu, Y. et al. (2023) 'Refining ChatGPT-Generated Code: Characterizing and Mitigating Code Quality Issues'. *arXiv*. Available at: http://arxiv.org/abs/2307.12596 (Accessed: 13 October 2023).

Lund, B.D. et al. (2023) 'ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing', *Journal of the Association for Information Science and Technology*, 74(5), pp. 570–581.

Lysova, E.I. et al. (2023) 'Ethics and the Future of Meaningful Work: Introduction to the Special Issue', *Journal of Business Ethics* [Preprint].

Makki, A. and Abid, M. (2017) 'Influence of Intrinsic and Extrinsic Motivation on Employee's Task Performance', *Studies in Asian Social Science*, 4(1), p. 38.

May, D.R., Gilson, R.L. and Harter, L.M. (2004) 'The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work', *Journal of Occupational and Organizational Psychology*, 77(1), pp. 11–37.

McKinsey (2022) *Mind the Gap*. Available at: https://www.mckinsey.com/~/media/mckinsey/email/genz/2023/04/2023-04-25b.html (Accessed: 15 November 2023).

Mortimer, S.A. (2023) 'What Makes Work Meaningful?', *Journal of Business Ethics*, 185, pp. 835–845.

Motowidlo, S.J. and Scotter, J.R.V. (1994) 'Evidence That Task Performance Should Be Distinguished from Contextual Performance', *Journal of Applied Psychology*, 79(4), pp. 475-480.

Newport, C. (2022) *The Year in Quiet Quitting*. Available at: https://www.newyorker.com/culture/2022-in-review/the-year-in-quiet-quitting (Accessed: 30 October 2023).

Northwest Executive Education (2023) *Top 8 ChatGPT Use Cases for Businesses*. Available at: https://northwest.education/insights/careers/top-8-chatgpt-use-cases-for-businesses/ (Accessed: 10 October 2023).

Noy, S. and Zhang, W. (2023) 'Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence', *SSRN Electronic Journal* [Preprint].

Nuñez-Varela, A.S. et al. (2017) 'Source code metrics: A systematic mapping study', *The Journal of Systems and Software*, 128, pp. 164–197.

OpenAI (2023) *Dall-E 3*. Available at: https://openai.com/dall-e-3 (Accessed: 07 November 2023).

Perkel, J.M. (2023) 'Six tips for better coding with ChatGPT', *Nature*, 618(7964), pp. 422–423.

Ramos-Villagrasa, P.J. et al. (2019) 'Assessing Job Performance Using Brief Self-Report Scales: The Case of the Individual Work Performance Questionnaire', *Revista de Psicología del Trabajo y de las Organizaciones*, 35(3), pp. 195–205.

Russell, S. and Norvig, P. (2021) *Artificial Intelligence*, Global Edition. 4th edn. Pearson Education.

Shank, D.B. et al. (2019) 'Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence', *Computers in Human Behavior*, 98, pp. 256–266.

Stegeman, M., Barendsen, E. and Smetsers, S. (2014) 'Towards an empirically validated model for assessment of code quality', in *Proceedings of the 14th Koli Calling International Conference on Computing Education Research*, pp. 99–108.

Surameery, N.M.S. and Shakor, M.Y. (2023) 'Use chat gpt to solve programming bugs', *International Journal of Information Technology & Computer Engineering (IJITC)*, 3(01), pp. 17–22.

Turel, O., Yuan, Y. and Connelly, C.E. (2008) 'In Justice We Trust: Predicting User Acceptance of E-Customer Services', *Journal of Management Information Systems*, 24(4), pp. 123–151.

Van Slyke, C., Johnson, R. and Sarabadani, J. (2023) 'Generative Artificial Intelligence in Information Systems Education: Challenges, Consequences, and Responses', *Communications of the Association for Information Systems*, 53(1), pp. 1–21.

Wilson, H.J. and Daugherty, P.R. (2018) *Collaborative Intelligence: Humans and AI Are Joining Forces*. Available at: https://www.forbes.com/sites/bernardmarr/2022/08/10/the-best-examples-of-human-and-robot-collaboration/ (Accessed: 10 October 2023).