

# The Impact of Generative AI on Human Productivity in Creative Writing

Vinay Bhimavarapu

Westfield High School

## ABSTRACT

With the innovative field of generative artificial intelligence having advanced at an incredibly rapid rate, its applications are of the utmost priority to study. The purpose of this study is to determine whether generative AI can help increase human productivity, specifically in writing. We hypothesized that generative AI will have a positive impact on human productivity. To test this, we employed the use of participants to write short fictional stories, one with the help of AI and one without. They were provided with survey questions that helped assess any changes in productivity levels. The productivity of the participants was also analyzed in terms of grammar, spelling, and consistency while compared against time. With the results obtained, we hoped to assess how AI can impact productivity in creative tasks (i.e., writing, art). We also hoped to understand its broader applications for human use and potential benefits and caveats to using generative AI. Based on the results, we concluded that using generative AI did indeed improve writing productivity as it lowered the number of errors and shortened the time taken. However, how productive the individual was in producing quality work of their own merit also depended on how much work they delegated to the generative AI as well as how they perceived it.

## Introduction

Will the advent of generative artificial intelligence bring us to a lower, more dependent state in which we begin to offload our critical thinking onto a computer, or will it elevate us to a higher level of power and efficiency? Since generative artificial intelligence is still a developing subject, its applications have only begun to be explored. For instance, one study, [1] *Generative Models in Artificial Intelligence and their Applications*, explored generative AI and its powerful ability to develop synthetic data, artificial data meant to mimic real data, which can fill information gaps to train models (Castelli and Manzoni, 2022). By using synthetic data, we can avoid privacy concerns stemming from real data. Another study, [2] *Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning*, discussed how maximizing the use of generative AI tools such as ChatGPT can also improve education, resulting in more personalized, interactive learning (Baidoo-Anu and Owusu Ansah, 2023). Companies such as Github have also capitalized on this new wave by releasing their own generative AI products such as Github Copilot. [3] In a study conducted by Github, Github Copilot was shown to improve programmers' perceived productivity as 88% reported they were more productive, 88% reported they completed their tasks faster, and 60% reported they felt more fulfilled (Github Next and Microsoft Office of the Chief Economist, 2021).

There is concern that this new technology could serve to replace humans and make many human jobs obsolete. People also worry that Generative AI will provide inaccurate data, carry pre-existing bias, and enable students to plagiarize, outweighing whatever benefits it has. However, under a more optimistic light, it can help people in their jobs by generating new ideas for them, automating menial subtasks, saving time, providing additional insight, and reducing the skill gap between workers. [4] In a study done by Nielson and Norman Group, participants used ChatGPT to aid them in writing business documents and the results for time spent during each process were compared with and

without AI. According to the data, the brainstorming and drafting phases were completed quicker but editing strangely enough, took longer.

The broader applications of generative AI have been explored in fields such as business, education, research, and programming through studies like those mentioned above. Overall, they discuss the general use of generative AI in developing and enhancing content as well as predicting and modifying data. However, the present study seeks to explore the applications of Generative AI in creative writing and overall, its application in aiding regular individuals in creative and intellectual tasks. The study will explore how the abilities of Generative AI mentioned above can work towards aiding people with its in-depth knowledge of fundamental structures of plot to supplement the user's work and build off of it by attempting to support it wherever it deviates from said structure, similar to how technologies like Github Copilot help support programmers.

Therefore, our study puts the applications listed in the previous studies to the test, assessing the effects of Generative AI on productivity in its use by everyday individuals in creative endeavors and generating and improving content. This study seeks to answer the question: What is the impact of Generative AI on human writing productivity?

## Method

### Demographics

In order to judge the impact of Generative AI on human productivity in writing, we recruited ten participants to each write one story by themselves and a second story with the help of ChatGPT. The participants are a combination of teenagers ranging from ages 14 to 18 as well as adults ranging from 30's to 40's to analyze any possible differences in the study in terms of age. The participants were tasked with writing short creative writing stories, each being a minimum of three paragraphs.

### Tasks

The first prompt for the story written without AI was *"Write a story about Saul the snake who wants to learn how to fly (6 sentences minimum)."* We designed this prompt to provide the participants with the opportunity to be imaginative so their independent writing proficiency and creativity can be better studied. The second story that was written with the help of ChatGPT tasks the participants with the prompt: *"Write a story about Jimmy the Frog, who is lost and can't find his pond (6 sentences minimum). Use AI to help you write or do this"*. This story was chosen as it is close to the same level of challenge as the first prompt while being similar in nature. The participants were tasked with using ChatGPT to help them write their stories as they saw fit and what they deemed necessary.

### Measures

#### *Qualitative Questions*

After writing each story, participants were asked to answer two open-ended questions including 1) what, if anything, they learned from ChatGPT, and 2) how they felt while writing the story. We asked this in an open-ended format in order to better understand the perspectives of the participants in their own words.

#### *Quantitative Questions*

Participants were also asked to answer quantitative Likert scale questions assessing 1) their perceived productivity, 2) perceived quality of the stories, and 3) personal satisfaction. By asking these questions, we hoped to easily compare

the differences in these feelings about their individually written story and their AI assisted story while also being able to find averages between the results of the participants. These questions helped us interpret the differences in how the participants felt overall. This also helped us determine how participants saw their productivity shift between using AI and writing on their own. The times in which each participant finished each of their stories was recorded in order to compare how fast they were able to write their stories with and without the use of AI. Additionally, they rated their writing proficiency to better understand the differences between their stories and compare them with other participants to account for any differences in quality and efficiency. They also described how many times they prompted ChatGPT for help, and how much of their story was written by the AI, expressed as a percentage. This helped us better understand the extent to which the AI contributed to their writing.

## Demographics

Demographic information such as age, hobbies, and writing proficiency were recorded to reduce unaccounted variables and potentially find covariates. To reduce any possible bias from the participants, not only did the study record their perceived quality of work but it also measured the quality of their work by assessing grammar, spelling, and consistency. The quality of their writing was then compared against the time they took to write in order to best assess their productivity.

## Results

### Human-written story

Upon collection of the results, we found the averages of all the quantitative measures and observed patterns between the qualitative responses while also observing word count, spelling mistakes, grammar mistakes, and plot consistency, the last of which was rated from 1 to 10 based on how coherent the plot was. For the solely human written stories, word count ranged from 61 to 445 ( $m=195$ ,  $SD = 134$ ). Spelling mistakes were generally low, ranging from 0 to 3 ( $m=1$ ,  $SD=1.134$ ). Grammar mistakes were more common ranging from 0 to 6 ( $m=2$ ,  $SD=2.999$ ). Inconsistencies ranged from 0 to 11 ( $m=3$ ,  $SD=4.489$ ). Time taken for solely human written stories ranged from 2 minutes and 3 seconds to 24 minutes and 6 seconds ( $m= 9$  minutes and 18 seconds,  $SD=0.003$  minutes). Quality ratings ranged from 4 to 7 ( $m=5.29$ ,  $SD=1.38$ ). Productivity ratings ranged from 1 to 7 ( $m=5.14$ ,  $SD=2.34$ ). Satisfaction ratings ranged from 4 to 7 ( $m= 5.86$ ,  $SD=1.21$ ). Reports on how the participants felt while writing the stories by themselves were generally favorable with the participants commonly indicating they felt engaged even if at first apprehensive with little negative opinion.

### AI-assisted story

As for the Artificial Intelligence assisted writing pieces, word count ranged from 87 to 392 ( $m=221$ ,  $SD=112$ ). Spelling mistakes ranged from 0 to 1 ( $m=0.28$ ,  $SD=0.49$ ). Grammar mistakes ranged from 0 to 7 ( $m=1$ ,  $SD=2.56$ ). The number of inconsistencies ranged from 0 to 3 ( $m=0.57$ ,  $SD=1.13$ ). Time taken to write the stories ranged from 3 seconds to 15 minutes ( $m= 4$  minutes and 57 seconds,  $SD=0.005$ ). The number of times the user prompted the chatbot ranged from 1 to 2 ( $m=1$ ,  $SD=0.45$ ). Quality of work ratings ranged from 5 to 7 ( $m=6.14$ ,  $SD=1.07$ ). Productivity ratings ranged from 1 to 7 ( $m=5.14$ ,  $SD=2.27$ ) being the average rating. Satisfaction ratings ranged from 3 to 7 ( $m=4.86$ ,  $SD=1.46$ ). When asked what, if anything they learned from ChatGPT, participants reported that they either learned nothing, learned how to write better stories, or that ChatGPT was simply good at writing stories, but they hadn't specifically gained any insight from it. When asked how they felt while writing the stories with the help of ChatGPT, participants generally reported feelings of boredom, ambivalence, with few expressing strong emotion. Percentages

of the stories written by ChatGPT ranged from 40% to 100% with the average percentage being 90%. Self-rating of user proficiency in writing ranged from 4 to 7 with 5.4 being the average value. The ages of the participants ranged from 14 to 48 with 21 being the average age. English grades ranged from B+ to A. Outside hobbies varied from reading books, to playing video games, as well as sports.

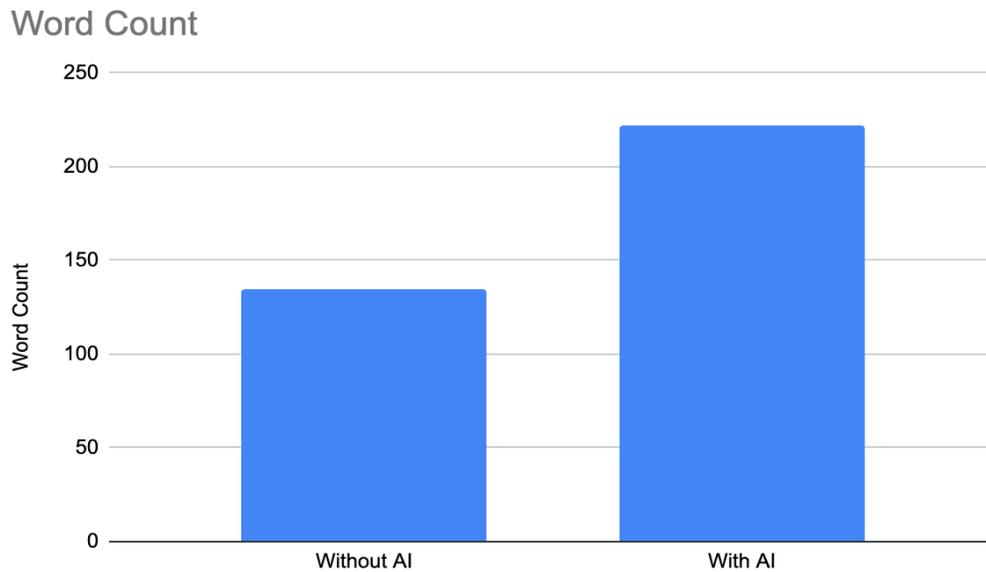


Figure 1. The differences in word count with and without the contribution of AI

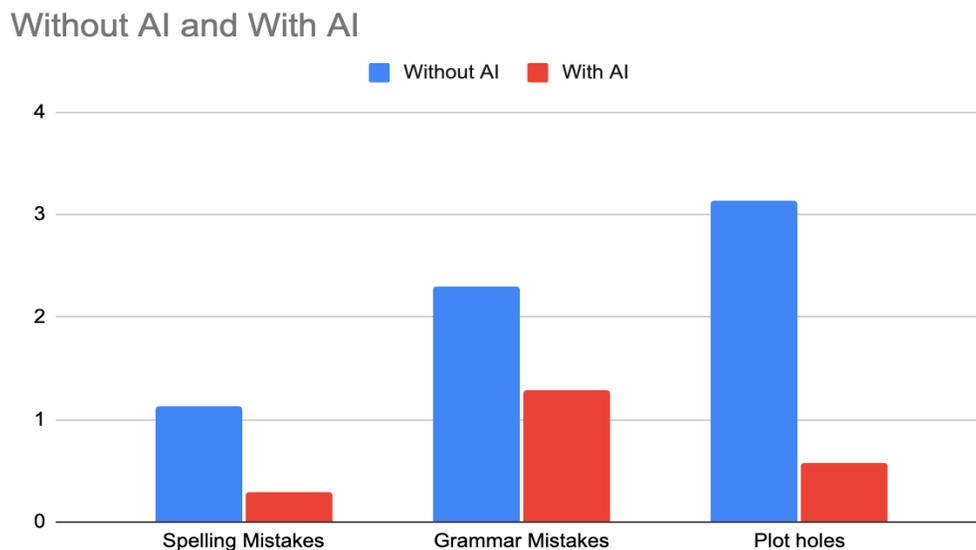
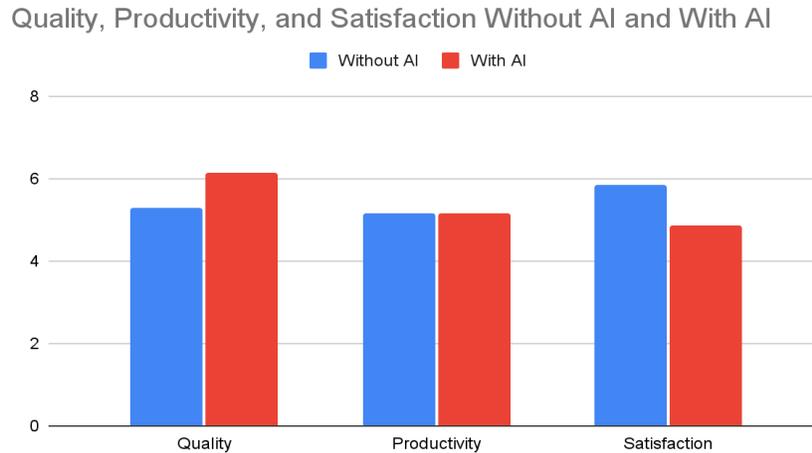
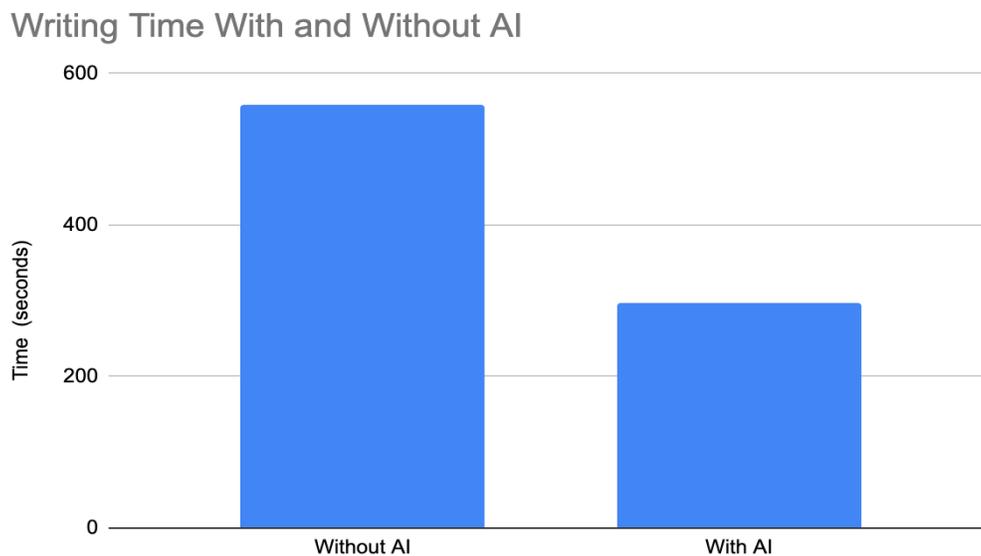


Figure 2. Comparing the number of spelling, grammar, and consistency errors with and without AI



**Figure 3.** Comparison of quality, productivity, and satisfaction with and without AI



**Figure 4.** Comparison of time taken to write story with and without AI.

## Discussion

### Errors and Consistency

Overall, the average word count increased from 195 to 221, with AI, showing that ChatGPT was able to help the participants produce more words on average with an increase of 13.3%. While the instances of spelling mistakes were low with or without AI, AI assisted stories had a smaller average amount of  $M=0.286$  compared to stories written without AI which had an average amount of 0.571 showing a decrease of 49.9%. Average grammar mistakes were also reduced from  $M=2.429$  to  $M=1.286$  showing a decrease of 89%. In terms of overall plot consistency, ChatGPT assisted writing was given a higher score of 9.2, compared to 7.4 without AI, for having more coherence and fewer

inconsistencies, showing an increase of 24%. Average time taken to write the stories was also shown to decrease with the use of A.I from M=00:09:18 to M= 00:04:57, showing a decrease of approximately 47%. As such, with the use of ChatGPT, writing has shown to increase in overall word count and plot consistency by a significant amount, decrease the number of average spelling and grammar mistakes by a marginal amount, and decrease the time taken by a significant amount of nearly 50% reduction.

## Variations Based on Participant Proficiency

Quality ratings for ChatGPT assisted stories did increase by one rating unit on average, but productivity ratings stayed mostly the same, while satisfaction ratings decreased by an average of one rating unit. Overall, more proficient writers were shown to have less of a difference in their writing quality whether they used AI while less proficient writers were shown to have a more significant boost in their writing quality. The proficiency of the writer was determined mostly based on their letter grade in English if they were a student or the level of writing quality, they exhibited in the stories they wrote by themselves. All participants took less time writing their stories with the use of AI. However, the increase in quality for writers that exhibited lower proficiency could be attributed to the greater contributions of ChatGPT to writing most or even all of the story rather than the merit of the writer themselves.

## Participant vs. AI Contributions

Stories were also shown to have more errors and inconsistencies the more the participants contributed to the story. Those who contributed little to nothing saw few errors and inconsistencies. However, those who had ChatGPT contributed less than or equal to 50% still saw fewer errors and greater coherence in their AI assisted stories. In terms of comparing the content of the writing pieces and the stories themselves, the AI assisted stories were shown to be more coherent on average. Some stories that were written solely by humans were quite poor in content, made little sense, and were not very engaging. However, others had levels of coherence comparable to the AI assisted stories and were also more engaging, exhibiting more personality and unique elements. In contrast, the AI assisted stories used more generic story elements and the more ChatGPT contributed, the less unique and engaging it became and the more recycled content there was.

For instance, one participant wrote about "Saul the Snake " being adopted by birds. In the story, he desperately wanted to fly because he was raised by birds but eventually realized he couldn't. However, his natural defenses as a snake meant he could still protect his family from threats. In their AI assisted story, in which they contributed 60%, they wrote about how Jimmy the Frog gets lost, consults a cricket who tells them where to go, and gets captured by a child who throws them in the exact direction of their pond. Despite the absurdity of the story, it still displays unique and engaging elements while also displaying fewer errors and being written in less time thanks to the help of ChatGPT. Another participant wrote a story about how Saul the Snake grew jealous of birds and their ability to fly. He deliberately hunted them down and eventually took a bird couple's eggs hostage so that they could carry him on their backs and simulate the experience of flying. However, the birds then kill him and feed him to their young. Despite the morbidity of the story, it still exhibits more unique elements and interesting characters than the AI assisted story which the participant only contributed 10%. In the AI assisted story, Jimmy the Frog gets lost, wanders around, and then finds his pond again and vows never to stray too far from it. Overall, the story was far less engaging and developed. Participants were shown to prompt ChatGPT very few times, either once or twice, mostly as they used the first prompt that ChatGPT gave them.

## Participant Perceptions of ChatGPT

In general, participants were also shown to have a mostly negative or ambivalent view of ChatGPT and agreed that they had not learned anything or merely learned about ChatGPT's abilities but gained no specific insights from them. The major consensus was that using ChatGPT was more convenient but less engaging and satisfying as it was not solely the work of the participants and was heavily influenced by AI. As one participant stated:

*"I was more satisfied with the first story than the second one because in the first story, I was able to use my thoughts and creativity. On the second one, I didn't feel productive, so that automatically made me feel a bit unsatisfied. Using Chat GPT made me feel like I haven't accomplished anything. It didn't feel like I worked hard or put in any effort. The first story that I wrote came from my heart and mind".*

This perspective could be explained by the idea that participants were more interested in their own writing than that of an AI and had a stronger connection to it. This also complements another theory regarding the results of the study that the participants' perceptions of the AI also hindered their ability to use it as a supplementary tool. Based on their responses, the participants perceived generative AI like ChatGPT as tools to do their work for them rather than something that could actively work with them and augment their abilities. For instance, one of the participants, a teacher, who would understandably have a natural inclination towards seeing ChatGPT as a tool to cheat rather than assist, prompted ChatGPT twice and let it write 100% of the story. As they did not perceive it as a tool to improve their own productivity or enhance their work, they simply let the AI write the entire story as the prompt only asked them to use ChatGPT however they thought necessary.

As such, like many technologies, generative AI has massive potential for good, being able to serve a symbiotic role for humans in enhancing their projects and working side by side with them. However, as those like the high school teacher have seen, humans abuse tools like ChatGPT to do their work for them rather than with them, getting their work done faster but not with their own merit. While the initial hypothesis is technically proven true, as participants were able to finish their stories faster with fewer errors, the full impact of generative AI on productivity depends on the percentage of contribution it makes to the given task.

## Significance of the Results in a Broader Context

Overall, generative AI is probably better used for more menial repetitive tasks that can allow people time for more fruitful pursuits rather than creative or intellectual tasks such as creative writing. These tasks require a uniquely human perspective, and they can still be useful for brainstorming, organizing ideas, and checking spelling and grammar. However, people may still rely too heavily on the abilities of AI to do their creative thinking for them because they view them as tools to work for them rather than with them. As with the example studies seen in the introduction section, tasks such as writing business documents and code might be better suited for input from generative AI as they require more human input, are not purely creative, and do involve repetitive tasks that can be automated to save time. However, the functions of AI can still be abused for tasks like these. Those who do know how to use generative AI to augment their abilities will have the opportunity to boost their productivity while those who use AI to evade their responsibilities will most likely fall behind. The impact that generative AI has on society depends on the perception each individual person has on the technology and how they use it.

## Limitations

Limitations in the study included 1) the minimum length required of the stories written, 2) varied interest levels of the participants, 3) the diversity in prompts used, 4) the diverse perspectives of ChatGPT. Because the prompts only requested that the participants write a minimum of six sentences, this meant that the AI would naturally make a larger

contribution. If the participants were tasked with writing longer, more full-length stories, it is possible that ChatGPT would make a smaller, more even contribution or even contribute less than the average participant for each story. Additionally, it is possible that the level of interest that the participants had in writing the stories or the prompts they were tasked with writing about, limited their productivity. If participants were able to write about topics that they found more interesting, they might have put more effort in the stories or contributed more than ChatGPT. The two prompts that the participants were tasked with writing about may have also differed in difficulty which may have affected their productivity levels differently. Additionally, the perceptions of ChatGPT may have affected the participants' ability to work with it, and as such, more diverse perspectives should have been included.

## Future Work

To further explore this research area, we could shift the different prompts between the same or different levels of difficulty, write the same prompt both with and without AI, write the AI assisted prompt first, or allow the participants more creative liberty to write about whatever they want rather than having to stick to a prompt.

Participants could also be compared based on proficiency levels, groups could be assembled comparing average people who have little interest in writing or are amateurs with more skilled or even professional writers to better assess the differences between AI assisted work between proficient and non-proficient writers.

Additionally, participants could be grouped based on their views of generative AI and if they have a strong positive, mild positive, neutral, mild negative or strong negative view of its capabilities in order to better assess how they use them. Participants could also read each other's work and rate it based on quality to better assess how much the writing changes from an unbiased reader perspective. A larger pool of participants would also be ideal to reduce sources of error. All of these measures could help build out the data found in this study and provide additional, more nuanced, insights.

## Conclusion

The initial hypothesis behind this experiment was that generative AI tools such as ChatGPT could help improve human productivity. While the AI assisted writing pieces had less errors and inconsistencies and took less time than the stories written solely by humans, the participants, for the most part, allowed ChatGPT to contribute quite heavily and write the great majority of their stories. This goes to show that the impact that generative AI has on human writing productivity as well as productivity in general in day-to-day life depends on how humans perceive the tool. Those who contributed more or equal to what ChatGPT did, saw improvements in their story which still carried their unique voice and creative thought. Those who had ChatGPT contributed most, or all of the stories did saw fewer errors but did not actually add to their own skills or make a story individual to them, but rather stories that were generic and recycled. Generative AI does have the capacity to improve average human productivity but only if they know how to use it to augment their own abilities rather than do their work for them. With the rate that generative AI is advancing, it seems that AI will be able to do many tasks that were once thought to only be able to be done by humans. Those who make a habit of relying too heavily on AI's generative abilities will find themselves without the necessary skills to survive in a world in which AI can surpass them. However, those who learn how to use generative AI to enhance and build off of their work will come to have a significant edge over those who do not. As such, it is irreconcilable based on the effect that generative AI has had on society and the many applications it has already been used for, the large role that it will come to have in the future. How people fare in this new era depends solely on them.

## References

- [1] Baidoo-Anu, David and Owusu Ansah, Leticia, Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning (January 25, 2023). Available at SSRN: <https://ssrn.com/abstract=4337484> or <http://dx.doi.org/10.2139/ssrn.4337484>
- [2] Castelli, Mauro, and Luca Manzoni. "Generative Models in Artificial Intelligence and their Applications." *applied sciences*, 2022, p. 3. SSRN, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4337484](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4337484). Accessed 2023.
- [3] GitHub Next and Microsoft Office of the Chief Economist. "Research: quantifying GitHub Copilot's impact on developer productivity and happiness." *The GitHub Blog*, 7 September 2022, <https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/>. Accessed 17 May 2023.
- [4] Nielsen Norman Group. *Time Spent On Writing Subtasks*. 2023. *LinkedIn*, [https://www.linkedin.com/posts/jakobnielsenphd\\_chatgpt-productivity-ai-activity-7051574746928320512-3DrA/?trk=public\\_profile\\_share\\_view](https://www.linkedin.com/posts/jakobnielsenphd_chatgpt-productivity-ai-activity-7051574746928320512-3DrA/?trk=public_profile_share_view).