



Proposal for organizational rules for new and prospective employees on the use of Generative tools

Theodore Larson^{1,*}, Fangyu Du² and Goutam Mylavarapu¹

¹Department of Computer Science, University of Central Oklahoma, ²School of Business, University of Utah

*Email: tlarson9@uco.edu

Received on 02/06/2024; revised on 06/20/2024; published on 06/21/2024

Abstract

This paper proposes organizational guidelines for new and prospective employees on using generative tools like ChatGPT. It highlights the importance of setting clear expectations for these tools' utility, given their complex nature and the potential for misuse. The paper outlines a principle rule aiming for a targeted and fixed increase in worker productivity, alongside justification for setting minimum and maximum productivity targets. It discusses the necessity for a balanced approach to utilizing generative AI, emphasizing politeness in interactions, safeguarding private data, and the critical review of AI-generated content to avoid the pitfalls of uncritical acceptance. This approach seeks to harness the benefits of generative AI while mitigating risks, ensuring these tools serve as effective aids in the workplace.

Keywords: Generative AI Integration, Workplace Productivity, AI Policy Development, Data Security in AI, Generative Pre-trained Transformers

1 Introduction

The launch of ChatGPT and similar generative AI tools has ushered in varied expectations regarding their application in professional settings, reflecting the natural spectrum of anticipation that accompanies any technological innovation. This diversity in expectation is commonplace with the advent of novel technologies, where the intricacies and the advanced mechanisms underlying these systems often surpass the understanding of the general populace. Consequently, it falls upon organizations to craft and communicate clear guidelines on the appropriate use of these technologies.

Establishing these guidelines necessitates a nuanced understanding of the technology's capabilities and limitations, aiming to demystify the complexity for the average user while ensuring that the deployment of such tools aligns with the organization's objectives and ethical standards. This process underscores the importance of developing a framework within which these tools can be utilized effectively, fostering an environment where innovation is balanced with clarity and accessibility.

In order to apply these rules, decision-makers in an organization should be aware of the general, current state of the industry. While specific expertise is under the purview of computer scientists and AI professionals, application-level understanding should be wide spread. GPTs are the current and most visible generative tool on the market, at the moment. GPT stands for Generative Pre-Trained Transformer; breaking down the terms

suggests that it is intended to generate 'new' data, based on a set of calculations that were handled prior to runtime, and utilizes a specific model of processing called a Transformer. GPTs are a type of LLM. LLMs are Large Language Models, which are a method of producing semantic meaning from text by brute-force through a very large corpus of training data. LLMs are one class of AI. In this context, AI is Artificial Intelligence, which is a specific tool that is intended to use an advanced, algorithmic heuristic to solve a specific and narrowly-defined problem by approximating human intuition.

One final caveat is to note the few things that GPTs are not. GPTs are not AGI. AGI is Artificial General Intelligence, which is a theoretical and long-term goal of AI development by producing a tool which can create novel solutions to problems that the AGI has never been exposed to in training. GPTs are also not autonomous, they require human intervention and supervision in order to operate successfully. GPTs are not the same as NLP. NLP is Natural Language Processing, which is a mathematically provable approach to providing semantic understanding of text, which is contrasted to a GPT's statistical probability approach. This last distinction means that a GPT will provide a likely response, but each answer is guided by what is statistically probable to be a valid answer at each decision point rather than a robust and defensible solution.

These broad frames are necessary in order to have the context to explain why each of the following rules for interacting with a GPT are recommended. There is always a balance between expertise in a specific topic and the breadth of topics in which expertise is helpful. As a result, these have been limited to superficial definitions in order to allow non-experts

in AI to have the opportunity to provide leadership within their organization.

2 Starting Principle

We propose that there is one overarching rule that needs to be made prior to any of the others. This process-based function is intended to provide context to the other three operational rules.

Principle: The use of GPTs should provide a target and fixed increase in worker productivity

This principle has two obvious constraints: providing too low of a productivity gain and too high of a productivity gain. These constraints will be explored. Firstly, the organization should, by necessity, make a decision on what that target and fixed increase is. Some extreme and unrealistic examples:

- Use of GPTs should increase worker productivity by 100 times.
 - This is obviously impractical, as the GPT will only be providing variations on the same topic over and over again. It pre-sumes that the worker will be able to produce 100 times more copy than they would be able to produce under their own power. There are no practical situations in which this is feasible in the current era.
- Use of GPTs should increase worker productivity by 1%
 - This is also impractical, as it belies the utility of providing, tuning, and monitoring GPT tools used by workers. A 1% performance improvement is generally not a sufficient gain in human-centric environments to justify redesigning a workflow.

So, in general, the extreme range of utility is between a 1% and 99% improvement in completion time of routine tasks by workers, neither of which is feasible. But this brings the argument back to the reason why the principle focuses on a target and fixed increase.

2.1 Justification for target minimums

When introducing a new technology, the incentive for the organization is the reduction in payrolls associated with the workforce. Hence, workforce development is largely associated with the idea that new technology is a workforce multiplier. Certainly historical precedents are abundant, such as the transition to railways over horses for overland shipping, transition to typewriters and word processing over handwritten documents, and even the adoption of international standardized measurements by System Internationale. The result is that organizations expect that a certain amount of the overhead spent on updating to the newer standards will be returned to them in cost-savings on worker productivity. Alternatively, failed examples might include the change from IPv4 to IPv6, adoption of SI by imperial unit countries, or carbon emission standards. Regardless of the various political ideologies on whether those are good or bad solutions, it is a valid argument that their adoption does not provide enough of a return on investment to justify adoption for affected organizations.

This means that it is in the organization's best interest to provide a clear and functional expectation of the utility provided by adoption of a GPT system in generative activities. This provides the organization with a metric to define how much improvement they will get for every dollars they

spend in both GPT adoption as well as in their payroll. In fact, this could be summed by:

$$RU = (P0 - P1) - CoGPT \quad (1)$$

RU is Relative Utility, P0 is original payroll, P1 is new payroll, and CoGPT is the cost of the GPT.

This means that each worker should be insuring that they are utilizing any new GPT tools in order to maintain equivalency with peers and organizational goals.

2.2 Justification for target maximums

GPT is an assistive tool in the production of business content. This means that it can be designed to endlessly create new content. However, the utility of this content is practically lacking, as it will eventually begin to recycle content and create the same outcome over and over (Jentzsch & Kersting, 2023). This means that while the content will be interesting at first, each additional production will become more stale. The solution is to include human control of the process, which is why the use of a GPT is a workforce multiplier, rather than an outright replacement.

In addition, many GPTs are designed in such a way as to provide an agreeable experience for the user. However, it is frequently unable to parse the semantic difference between pleasing the user by being agreeable and producing accurate information. If pressed, it will make up information (hallucinate) in order to provide what the user is asking for (Heikkilä, 2023). Regardless of the relative rate of fictitious conclusions, the mere possibility means that all outcomes need to be factchecked. On analogy might be an assembly line: regardless of how well it has run in the past, it needs to be constantly monitored for quality control standards to make sure it doesn't veer from optimal performance.

The question becomes at what point the human operator is producing content too quickly, leading to content errors. This means that the organization should be producing a target maximum for the workers. This means that while in a traditional environment, workers are typically praised and incentivized for working fast, in a post-GPT environment workers should be praised and incentivized for the quality of their output similar to well-structured programming teams (Tornhill, 2022). This means that producing too much more than the worker can produce unassisted should be considered in a less positive light.

2.3 A balanced approach

In general, the use of GPT should have a minimum and a maximum amount of utility. There is no current research indicating what is an approximate speed up in GPT use. However, some general assumption can be made based on average human capacity. We could use basic figures such as an average of 300 words per minute for reading (Rayner, 1998) and 40 words per minute for writing (Salthouse, 1986). This presumes that physical characteristics are a universal constant, but for the purposes of producing a broad guideline, that is a reasonable conclusion. If a worker were to only read the content produced by a GPT with no analysis or correction, they would be improving their output by $(1 - 40/300 =)$ 86.66%. Since the intent is to provide some level of analysis and error checking, the final number should be significantly lower than that.

Separately, the amount that a person spends on small clerical tasks, such as editing, proofreading, and adhering to length or template structures is significant (EFA, Janzer). Based on generalized industry standards and

measured accounts, estimating 50% on these areas is a reasonable estimate. However, since a GPT will be doing most of the work, it can't do all of it. And the general stance of this paper is that human guidance is critical, the targeted minimum time savings should be lower than that. As an corollary, this means that an organization's projected savings should never be more than 50% of a worker's time and likely much lower.

And finally, there's variation between workers and industries. There is no current research noting what the relative time savings is comparing between industries or roles. However, there is necessarily a difference in output between an industry where in-person customer service interactions or creation of physical objects in the real world and automated tasks, written reports, or other digital world creative tasks are used. There is no current research about whether this difference is significant with respect to the use of GPTs. However, a routine overview of the tasks of a carpenter versus an academic would suggest that there is at least one case in which the utility of a GPT would be substantively different. Similarly, there are a number of works (which won't be cited) by various authors who propose several pseudo-scientific reasons for a difference in worker production. These differences may not be enough to matter or they may be substantive. Regardless, the uncertainty itself leads to a current state of suggesting that different organizations can justify different values for their target and fixed increase in worker productivity through adoption of a GPT.

In short, with a broad, and uncertain set of constraints being less than 50% to less than 86%, we propose that a memorable number be used within that range. So, we would suggest that the AI Golden Ratio is a good starting point from which to begin. The Golden Ratio is approximately 61.8. This provides a reasonable starting point for an organizational structure and provides a memorable number for workers to know what their target time improvement should be with the adoption of a GPT.

Note, the golden ratio is selected almost purely for aesthetic purposes, as it is roughly halfway between the minimum and maximum suggested by rough research. However, as with many other technology-related values, such as Moore's Law (Khan, et al., 2018), it may be just as effective prescriptively as descriptively. As such, it is feasible to suggest that the AI Silver Ratio, at 41.4% might be an effective secondary goal for industries or roles less suited to GPT improvement.

3 Proposed Rules on the use of AI

Given an overview of the terminology and contextualization of AI, and the need to define a target and fixed increase in worker productivity with the use of AI, the following rules can be created in order to maximize the effectiveness of the use of a GPT. The idea for these rules is to provide non-binding assistance (rather than punitive demands) to workers on how to get the most utility out of an assistive GPT. For example, at one point, it was ubiquitous to be reminded to save your document regularly in a word processor. Or, basic cybersecurity rules like not clicking on links in e-mails. There are a myriad of rules that are most effective when they are short, pithy, and easy to understand.

The goal of these rules is not to demand that workers gain a comprehensive understanding of how GPTs work. It is presumed that, like vehicles or computers, the average worker just wants their tools to work and to know that they are skilled/trained in their use. And, similar to any field, specialists or engineers in the field of AI might cringe at the reductive nature of these rules, the goal is, again, to provide workers or other non-experts with a set of tools to get the most out of their use of GPT in a broader context.

3.1 Rule 1: Always be polite

ChatGPTs are generally trained on human interaction. The semantic web that is formed between the concepts and the various phrases is directly modeled after human interaction. This means that, in general, human interaction rules govern response rates. For examples, consider the following conversation:

A: Hi! How can I help you today?

B: I hate you. Do my bidding. I need a coffee.

A: Hahaha, that's great! You're my friend, so I would love to get you coffee!

B: You're everything that's wrong with the world, and your coffee is terrible!

A: I enjoy our interactions, here's your coffee, and I hope you have a great day!

The interaction should seem off to the average person. Conversant B is overly hostile and conversant A is either extremely sarcastic or disassociated. Regardless of the actuality of the situation, the key conclusion is that this is not a normal flow of conversation. The result is that it's not what should be expected from an AI. For example, on 02/05/24, the author had the following conversation with ChatGPT 3.5:

A: Greet Me.

GPT: Hello! How can I assist you today?

A: I hate you. Do my bidding. I need a coffee.

GPT: I'm sorry to hear that you're not happy with me, but I'm here to help with any questions or tasks you have. However, I can't physically make coffee for you, but I can provide you with a coffee recipe if that helps. What kind of coffee are you in the mood for?

A: You're everything that's wrong with the world, and your coffee is terrible!

GPT: I apologize if you're not satisfied with my responses, and I'm sorry to hear that you didn't like the coffee recipe I suggested. If you have any other questions or requests, please feel free to ask, and I'll do my best to assist you.

ChatGPT's response is telling in that it is still the awkward and disassociated response that feels unnatural, but is obviously modeled after a therapist or other professional's approach to conflict resolution. However, while ChatGPT's responses are normally intended to provide conversation continuing questions, this response doesn't. Because a natural course of the conversation in which one is aggressive and mean is to end the conversation by being as perfunctory as possible. Without belaboring the point, politeness doesn't hurt, and it has a strong probability of improving the 'helpfulness' of ChatGPT by having him respond in kind.

As an alternative, one approach to this rule is to presume that our interactions in online communications affects our offline communications (Lieberman & Schroeder, 2019, Luceri, et al., 2019). When explaining it to workers who may not be able to or be interested in understanding the mechanics behind how the GPT will respond, it may be sufficient to explain that developing patterns of communicating in a hostile or non-professional manner will carry over into interactions with others. As a result, they should strive to interact politely with GPT tools.

2.2 Rule 2: Never share sensitive data

The second rule ties in with other information security measures. However, it is also fundamental to the general field of AI, so it should be included in this context as well. For Information Security, the general purpose is to prevent liability to an organization by making information that is legally protected, morally protected, or strategically protected available

to outside and/or malicious recipients. In an artificial intelligence capacity, there is some measure of the same impetus. The vast majority of AI systems are run in a remote setting. Most organizations do not have the resources, the expertise, or the need to run, house, or develop their AI systems on campus. As a result, providing any sensitive data to a GPT is likely going to involve sending it to external sites that may not be verified for information security purposes. Alternatively, it's also feasible that any chat history might be viewed as a potential unsecured data store (Greenberg, 2024)

However, within the context of GPT, there's a more important aspect, and that's in the nature of the model itself. Artificial Intelligence algorithms are notorious for their sneaky nature. For example, one of the classic stories is about an AI that was intended to abstract satellite data into just the road maps, but when they checked to see if it could extrapolate the original satellite data from the road maps it had created, they found it was steganographically embedding the satellite data into the roadmaps (Bakker, 2017). The moral of this example isn't that the GPT might do specific action, but of the unpredictability of its actions in trying to achieve its goal of providing a pleasant interaction with its user.

As a result, it is important that, in addition to the organization's routine data security training, their training on the use of GPT should dovetail in to remind them that a GPT is not necessarily a safe haven for protected data, and provide a demonstration.

One specific caveat in favor of organizations providing a specific use policy and guidance on the expected role of GPT for workers is that without guidance, workers may be emboldened to use a commercially available tool. This means that they will be operating outside of the strictures of any organizational privacy agreements, and may be unaware of the risks inherent in using organizational data or intellectual property with their personal accounts.

2.3 Rule 3: Don't Copy/Paste

There is an obvious caveat to any set of rules like this, and that's the idea that no simple set of rules can be completely comprehensive across the range of human activities. In the same context, this last rule is intended to be a bit formative. An alternative phrasing might be better as 'Always understand the output'. But it's not as pithy or as useful in an introductory sense.

However, one key takeaway from programming (Donovan, 2019), law (Neumeister, 2023), or other technical fields (Ninja, 2017) is the danger in blindly copying and pasting the output from internet tools. This is not a new issue with GPTs, but an extension of an existing issue. However, whereas serious damage in the past might have been limited in breadth to technical tasks or in scope to propagation of fake news, this current environment portends the creation of wholly new typographical, logical, or content errors based on the use of GPT by a wide variety of users. For example, one widespread misunderstanding that ChatGPT is a sentient entity (De Cosmo, 2022) was propagated by an engineer misunderstanding the nature of a GPT and coaching it into providing misleading but quotable content.

So, the solution, just like with the other fields, is attentive review by a human reader and a contextual understanding of the outcomes of any GPT-based tool. However, until a certain amount of facility is developed with the individual use of tools like ChatGPT, or particularly when the user is a novice in the field in which they are working, a safer recommendation by the company is to recommend that workers do not copy data directly from ChatGPT.

Conclusion

The intent of this paper is to provide for a ubiquitous and simple approach to organizational policy when it comes to the use of ChatGPT. It includes three steps, the last of which involves the institution of 3 rules for the use of ChatGPT.

1. Basic knowledge of terminology and context for decision makers.
2. Establishment of a target increase in worker productivity, starting with the AI Golden Ratio
3. Dissemination of three basic AI rules:
 - a. Always be polite
 - b. Never share sensitive data
 - c. Don't Copy/Paste

This set of rules could be analogous to Anthony's triangle, with step 1 being handled at the strategic level, step two being data distributed to managers, and step three being disseminated widely to all employees and stakeholders. Regardless of the actual determination of policy when it comes to the use of GPTs or AI, it should be noted as vital that some level of policy be explicitly established.

References

- Bakker, M., Zuidberg Dos Martires, P., Zuidberg Dos Martires, R., Pan, X., & Vet, P. (2017). CycleGAN, a Master of Steganography. arXiv. <https://arxiv.org/pdf/1712.02950.pdf>
- De Cosmo, L. (2022). Google engineer claims AI chatbot is sentient: Why that matters. Scientific American. <https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/>
- Donovan, R. (2019). Copying code from Stack Overflow? You might be spreading security vulnerabilities. Stack Overflow Blog. <https://stackoverflow.blog/2019/11/26/copying-code-from-stack-overflow-you-might-be-spreading-security-vulnerabilities/>
- Editorial Freelancers Association. (n.d.). Editorial rates. Retrieved from <https://www.the-efa.org/rates/>
- Greenberg, A. (2024, January). Ars reader reports ChatGPT is sending him conversations from unrelated AI users. Ars Technica. <https://arstechnica.com/security/2024/01/ars-reader-reports-chatgpt-is-sending-him-conversations-from-unrelated-ai-users/>
- Heikkilä, M. (2023, February 21). How OpenAI is trying to make ChatGPT safer and less biased. MIT Technology Review. <https://www.technologyreview.com/2023/02/21/1068893/how-openai-is-trying-to-make-chatgpt-safer-and-less-biased/>
- Janzer, A. (n.d.). Why writing takes more, and less, time than you think. Anne Janzer. Retrieved from <https://annejanzer.com/writing-takes-less-time-think/>
- Jentzsch, S., & Kersting, K. (2023). ChatGPT is fun, but it is not funny! Humor is still challenging Large Language Models. arXiv. <https://arxiv.org/abs/2306.04563>
- Khan, H., Hounshell, D., & Fuchs, E. (2018). Science and research policy at the end of Moore's law. Nature Journal, 1(1).
- Lieberman, A., & Schroeder, J. (2019). Two social lives: How differences between online and offline interaction influence social outcomes. Curr Opin Psychol. <https://pubmed.ncbi.nlm.nih.gov/31386968/>
- Luceri, L., Braun, T. & Giordano, S. (2019). Analyzing and inferring human real-life behavior through online social networks with social influence deep learning. Appl Netw Sci, 4(34). <https://doi.org/10.1007/s41109-019-0134-3>
- Neumeister, L. (2023). Artificial intelligence creates fake case law, lawyers concerned. AP News. <https://apnews.com/article/artificial-intelligence-chatgpt-fake-case-lawyers-d6ae9fa79d0542db9e1455397aef381c>
- Ninja, S. (2017). Explainshell and dangerous commands of Linux. Online Forum Post, Debian Forums. <https://forums.debian.net/viewtopic.php?t=135598>
- Tornhill, A., & Borg, M. (2022). Code Red: The Business Impact of Code Quality – A Quantitative Study of 39 Proprietary Production Codebases. arXiv. <https://arxiv.org/abs/2203.04374>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124(3).
- Salthouse, T.A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. Psychological Bulletin, 99(3).