Can ChatGPT Satisfy All? An Experimental Evidence

Preeti Sharma[1], Sourav Banerjee[2] and Anupam Bhattacharya[3]

[1] University of Engineering and Management, Jaipur, India
[2]University of Engineering and Management, Jaipur, India
[3] Institute of Engineering and Management, Kolkata, India
preeti.sharma@uem.edu.in

Abstract. Generative artificial intelligence (AI) is still in its nascent stage. With rapid advancements in terms of technology and use, researchers hold diverse views about its ability to complement natural intelligence. Out of several apprehensions regarding its adaptability, the aspect of self-satisfaction is yet to be empirically studied. In this paper, we demonstrate a set of experiments on two groups of postgraduate students. Each group contained students with and without expertise on ChatGPT. Each student in both the groups are given tasks to develop two technical reports: with and without using generative AI tools respectively. The first report was generic in nature, whereas the second report was more domain-oriented requiring deeper understanding and complex search. The outputs are measured on three aspects: report generation time, self-satisfaction and search utility. Students with AI expertise took significantly more time for domain-specific topic under ChatGPT. Ownership and associated pride were significantly higher in self-generated reports. For experts, AI-generated reports for generic topic showed more enrichment. Ownership and pride is found to be higher in AI-generated domain-reports when compared with that of AI-generated generic reports amongst the expert group. The second group is controlled over incentivised mechanism, and they also underwent a short AI training program. The incentivised group demonstrated significantly higher ownership, pride and enrichment along with marginally lesser standard deviation across all variables. Positive correlation is observed between search time and satisfaction amongst the expert group. Based on the final result, a theoretical framework of 'Natural-Artificial Indifference-curve' is proposed for further experiment.

Keywords: Generative AI, ChatGPT, User satisfaction, Indifference curve.

1 Introduction

Since its inception in November 2022, ChatGPT has become the fastest growing user application, crossing 180 million users.[1] According to USB analysts "In 20 years following the internet space, we cannot recall a faster ramp in a consumer internet app"[2]. Advanced artificial intelligence (AI) system handholds the generative AI application, simulating human brain for daily work requiring large data (Bengio, Lecun & Hinton, 2021). The application is set to access publicly available information across the web, like a human would do for creating essays, images, presentations and the like, commonly perceived as 'creative content' (Bin-Nashwan, Sadallah & Bouteraa, 2023); just that the efficiency is manifold in terms of speed and accuracy. For instance, an experimental study in the University of Minnesota, ChatGPT demonstrated the capacity of moderately earning a university degree (Choi et al., 2023). Moreover, ChatGPT, along with other advanced generative AI tools, is capable of providing high synergistic effect, where diversified knowledge pool over vast training data can collaborate in terms of higher productivity (Rice et al., 2024) juxtaposing lower cost (Hoffman et al., 2018).

But there are multiple facets of this generative AI, over which researchers are showing significant concerns. Firstly, of course, is the apprehension of routine job replacements and up-skilling requirements (George, George & Martin, 2023); secondly, the ethical considerations that seem to suffer reconcilable bias (see Stahl & Eke, 2024), and even over-reliance and subsequent decrease in cognitive ability (Bai, Liu & Su, 2023). Noy & Zhang (2023) addressed certain user-level concerns such as whether such tools would increase or decrease the skill inequality; whether it will complement or substitute tasks; and very interestingly, will it affect user satisfaction at large. In this paper, we focus on this last issue at an experimental level.

There are two verifiably different dimensions that can be attributed to worker dissatisfaction due to technological change. The first is the inherent fear of adaptability that is detrimental to productivity

P. Sharma et al. (eds.), *Proceedings of the International Conference on Artificial Intelligence in Management for Business and Industrial Growth (AIMBIG 2025)*, Advances in Economics, Business and Management Research 355,
https://doi.org/10.2991/978-94-6463-898-1_10

(Schwabe & Castellacci, 2020). Noy & Zhang (2023) experimentally validate this concern. Several models and frameworks of technology acceptance have been developed and widely used by researchers. The technology acceptance model (TAM), unified theory of acceptance and use of technology (UTAUT), AI device use acceptance model (AIDUA), etc.[3] However, the second dimension of dissatisfaction that emerges from the inherent ability of the user has not been researched yet. Ever since ChatGPT was made widely available to all for free, people with no technical background could also feel connected to AI and fulfil customised requirements (Wang, Lin & Shao, 2022). However, the question lingers on about whether the acceptability is all pervasive or not.

It is apparently understood that essays, reports, articles, etc., and creative writing such as stories, poetries, etc., have always had two categorically different human touch (Kaur & Saini, 2014). But the common aspect is that the writer takes the ownership and the pride associated with the creation. They are derived directly from the *perceived differences* in the creation (Brown, Pierce & Crossley, 2014). Then, the amount of search that goes into creating the material also involves satisfaction (see Huuderman, Wilson & Kamps, 2016) and knowledge enhancement. It is also true that information overload can easily offset and nullify the utility derived from the specific search that often leads to misjudgement of the choice set; still, in the present context, generative AI completely removes the effort of information search; the knowledge enhancement and the utility associated with it. The AI-generated creation apparently would also not have the ownership and pride associated with any kind of creation—commodity, service, or content. It has also been experimentally proved that more writing actually enhances writing ability (Fernandes, 2011). So, heavy reliance on generative AI may be derogatory in terms of writers' enrichment. Hence, in this study, we explore this issue from the user's perspective. The objective of the research is to judge the extent of satisfaction (or utility) the user derives from contents generated by ChatGPT in terms of ownership, pride, knowledge enrichment and information search, when compared with self-generated contents. The scope is limited to the psychological appeal to the user alone, and not on long-term impacts of AI reliance.

2 Methods

The experiment was conducted over postgraduate students of premiere management institutes in Kolkata and Jaipur, India. The study was conducted at two levels. In the first level, 140 students from UEM Jaipur were segregated in two panels A and B. Panel A comprised 94 students having high expertise in using generative AI and Panel B comprised of 46 students with little or no expertise. Each student was asked to generate a report each on two topics through ChatGPT. The first topic was generic in nature, requiring more of publicly available information: ***India's response to Covid 19 crisis.*** The second topic was domain-specific requiring higher depth and specificity that in turn necessitates complex search: ***Ukraine war and its impact on international trade.*** The same pair of topics were administered again, and the participants were asked to develop the report without the use of generative AI. The participants were, however, allowed to access the internet without any restriction. While using ChatGPT, the participants were also allowed to interfere with the generated content or even regenerate multiple times to reach to the desired satisfaction.

After the experiment, the participants were asked to fill out a response form that contained questions pertaining to self-satisfaction with reference to ownership, pride, knowledge enrichment, information search and AI expertise (high or low). Other than the foresaid responses, the ChatGPT generation time and its corresponding information search time of self-generated reports were recorded over the course of the experiment for each participant. Each was observed over a stopwatch from the time they began the search till they completed the report.

In the second group, 92 students were from Kolkata, India and the study was conducted in National University of Singapore (NUS). Panel C had 59 students of high expertise, and Panel D had 33 students with low AI expertise. The study was controlled over an inventive mechanism. The group of students was given course completion assignments as a part of their study abroad program in NUS. The experiment was put up as a course task in order to get the course-completion certificate. Furthermore, the students received 1-day training on the use of generative AI during their stay in Singapore. Hence, we could tap the shift in responses, as well as the performances of the students who were initially not well versed with ChatGPT.

AI tools gets positively affected by short-term incentives on the output generated. A study conducted by Noy and Zhang 2023 showed that when employees were exposed to generative AI under organizational incentive, they were not only more likely to embrace the opportunity; they were even more prone to repeating the same in the future. Therefore, the same sequence of experiments was carried out in NUS. The generic topic given to them was: *How you perceive Singapore when compared with your native country*. The domain-specific topic was: *Paradigm shift in sustainable business*. Table 1(a) provides the descriptive statistics for panel A, B, C and D over self and AI-generated reports, for generic and domain topics. The post-training statistics of Panel D is provided in Table 1(b).

In the final experiment, Table 1(b), the students were compared only on the domain topic that were administered once more. The generic topic was not considered due to the experimental *bias* we encountered. We shall elaborate the *bias* in Results section.

Table 1(a). Self and AI generated report statistics for Generic and Domain topics.

	Generation time		Ownership		Pride		Enrichment	
	Own	ChatGPT	Own	ChatGPT	Own	ChatGPT	Own	ChatGPT
	Generic/Domain	Generic/Domain	Generic/Domain	Generic/Domain	Generic/Domain	Generic/Domain	Generic/Domain	Generic/Domain
Panel A								
Observations	94	94	94	94	94	94	94	94
Mean	30.89 / 42.47	7.89 / 12.8	3.93 / 4.31	2.37 3.62	4.49 / 4.01	2.71 / 3.54	3.5 / 4.31	3.88 / 4.25
(S.D.)	(6.05 / 7.23)	(3.58 / 3.61)	(0.8 / 0.55)	(1.16 / 1.27)	(0.73/0.96)	(1.47/1.26)	(0.94/0.69)	(1.31/0.75)
Control	No	No	No	No	No	No	No	No
Panel B								
Observations	46	46	46	46	46	46	46	46
Mean	30.52 / 39.5	11.2 / 8.1	4.1 / 4.58	4.56 / 2.3	3.73 / 4.32	4.65 / 3.1	3.84 / 3.95	2.17 / 2.6
(S.D.)	(7.3 / 7.3)	(3.87 / 3.82)	(1.2 / 0.65)	(0.83 / 0.81)	(0.93 / 0.79)	(0.6 / 1.2)	(0.82 / 0.82)	(1.18 / 1.16)
Control	No	No	No	No	No	No	No	No
Panel C								
Observation	59	59	59	59	59	59	59	59
Mean	35.6 / 40.32	5.59 / 5.98	3.95 / 4.38	2.89 / 4.1	4.23 / 4.32	2.88 / 3.92	3.90 / 4.29	4.06 / 2.36
(S.D.)	(5.58 / 8.89)	(2.57 / 1.94)	(0.98 / 0.3)	(0.84 / 1.02)	(0.68 / 0.89)	(1.59 / 1.24)	(0.98 / 0.72)	(1.02 / 1.19)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel D –								
Observation	33	33	33	33	33	33	33	33
Mean	39.8 / 42.7	3.32 / 2.89	4.22 / 4.59	4.38 / 4.01	3.92 / 4.28	4.21 /3.99	4.28 /4.57	2.89 / 2.89
(S.D.)	(7.9 / 9.2)	(0.84 / 0.56)	(0.8 / 0.54)	(0.77 / 0.52)	(0.76 / 0.4)	(0.52 / 0.51)	(0.57 / 0.44)	(0.50 / 0.41)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 1(b). Self and AI generated report statistics for Generic and Domain topics.

	Generation time		Ownership		Pride		Enrichment	
	Pre-training	Post-training	Pre-training	Post-training	Pre-training	Post-training	Pre-training	Post-training
Observations	33	33	33	33	33	33	33	33
Mean	2.89	2.54	4.01	3.68	3.99	3.69	2.89	2.94
(S.D.)	(0.56)	(0.41)	(0.52)	(0.47)	(0.51)	(0.4)	(0.41)	(0.54)

Table 2(a). Paired *t*-test statistics

Pairs	*t*-value (95%)	*t*-value (Control)	Std. Error	Std. Error (Control)	DOF	DOF (Control)	Sig. Level	Sig. Level (Control)
Expert_Own_Self_Gen - Expert_Own_AI_Gen	9.613	2.50	0.16	0.21	93	58	.000	.015
Expert_Pride_Self_Gen - Expert_Pride_AI_Gen	10.63	9.04	0.17	0.21	93	58	.000	.000
Novice_Own_Self_Gen - Novice_Own_AI_Gen	-3.23	-2.42	0.14	0.15	45	32	.002	.001
Novice_Pride_Self_Gen - Novice_Pride_AI_Gen	-6.28	-3.77	0.15	0.13	45	32	.000	.000
Expert_Own_Self_Domain - Expert_Own_AI_Domain	5.11	3.55	0.13	0.20	93	58	.000	.001
Expert_Pride_Self_Domain - Expert_Pride_AI_Domain	3.02	2.58	0.15	0.20	93	58	.003	.001
Novice_Own_Self_Domain - Novice_Own_AI_Domain	14.78	4.20	0.15	0.25	45	32	.000	.000
Novice_Pride_Self_Domain - Novice_Pride_AI_Domain	8.01	3.88	0.15	0.21	45	32	.000	.001

Output generated through IBM SPSS (ver. 23)

Table 2(b). Correlation Statistics

Pairs	Pearson's Coefficient	Sig. Level	Control
[Panel_A_Generic_AI_Time]-[Panel_A_Own_AI_Generic]	0.569	0.01	No
[Panel_A_Domain_AI_Time]-[Panel_A_Own_AI_Domain]	0.584	0.01	No
[Panel_A_Generic_AI_Time]-[Panel_A_Pride_AI_Generic]	0.382	0.03	No
[Panel_A_Domain_AI_Time]-[Panel_A_Pride_AI_Domain]	0.601	0.00	No
[Panel_C_Generic_AI_Time]-[Panel_C_Own_AI_Generic]	0.494	0.01	Yes
[Panel_C_Domain_AI_Time]-[Panel_C_Own_AI_Domain]	0.644	0.01	Yes
[Panel_C_Generic_AI_Time]-[Panel_C_Pride_AI_Generic]	0.589	0.00	Yes
[Panel_C_Domain_AI_Time]-[Panel_C_Pride_AI_Domain]	0.611	0.01	Yes

Note: Panel B and Panel D are omitted for Correlation observations because of possible bias, non-coherence and unfamiliarity with the search process. Hence, higher search-time is not an indicator for better results.

3 Results and Discussion

We decipher the results from both the panels and the controlled groups separately. 'Ownership' and 'pride' associated with self-generated reports have consistently higher average when compared with that of AI-generated reports (Table 1a). SPSS (v. 23) output of paired *t*-test values are presented in Table 2(a). There is one exception where, non-experts while generating ‘generic’ topics felt higher ownership and pride (differentiated by negative *t*-values of -3.22 and -6.28, respectively). It is evident that this group of observation suffered from a psychological bias, because they embraced a new technology for the first time, which made them get overwhelmed by the sheer experience of text generation by ChatGPT. But when they were asked to generate a domain-specific report again, the results were completely different. The same bias was also observed within the incentivised groups. It is to be observed that the entire group had generated several types of contents through generative AI by this time (over and above the observed experiment), hence the novelty factor causing the foresaid bias was significantly reduced. Panel D, in generic topic, represented in Table 1(a) shows significantly high ownership and pride when compared with domain-specific topic. In Table 2(a), we find much less ownership and pride among non-expert students (validated by the highest *t*-value of 14.78, and third highest *t*-value of 8.01, respectively).

Pertaining to the foresaid results, two other important connections hold. Firstly, the time dedicated towards generation of report, irrespective of the source and the group, and the level of ownership and pride, seemed to have a significantly high positive correlation (Table 2b). This validates the importance of dedication and seriousness amongst the students. Hence, in descriptive statistics in Table 1(a) and (b), Panel A and B, we find much higher standard deviation.

When compared with panel C and D—the controlled groups. Secondly, we observe much higher self-satisfaction over all three variables—ownership, pride and knowledge enrichment in the controlled groups, especially within the “expert” category. In this context, ‘generic search’ by panel C showed much higher knowledge-enrichment (2.36 as against 4.06). The same panel, while going for domain-specific complex search, showed significantly less enrichment (2.36 as against 4.06). ChatGPT could not outperform ‘natural search’ in terms of objective and complex knowledge generation. Here, we clearly observe that Panel B recorded almost similar enrichment (mean) and similar within-group error (S.D.) over generic and specific topic. This similarity stems from the argument that the said non-experts could not distinguish between generic and complex search and the outcomes therefrom, over the trained data of generative AI.

This argument is also validated from another interesting observation. AI experts took more time in generating ‘domain-specific’ reports, whereas non-experts took lesser time in the same (Panel A against Panel B, and Panel C against Panel D in Table 1a). They were clearly unaware of the search complexities associated with complex topics. In the context of search time, experts and non-experts were seen to dedicate time for completely different reasons. Since the participants were given the liberty to generate the report multiple times, we observed that non-experts generated similar reports over multiple runs, whereas experts tried to refine the generation process with newer keywords.

Data from Panel D, elaborated in Table 1(b), suggested interesting insights into the effect of Training Program. We observed generalised decline in search time, ownership and pride. But the participants felt more enriched post-training. It establishes that fact that with higher expertise in the nuances of the search-process, users are able to generate more informative reports through ChatGPT.

The final set of questions in the post-experiment survey relates to the perceived impact of the additional information-search on the self-generated reports. Two questions were addressed to the expert-incentivised groups: *the degree of additional information search needed in self- generation*, and *the impact of the additional search in the self-generated content*. When it comes to browsing utility, existing metrics aim at capturing the quality of user interaction with the search engine results (Yilmaz et al., 2010). Furthermore, indifference curves have been seen to follow (or travel) with the consumer choice set (Drolet, Simonson, & Tversky, 2000). And choice sets, in terms of learning is directly related to incentives (Grove & Wasserman, 2006). Hence, it was only apt to keep the utility survey restricted to the incentivised group, who knew how to interact with the generated results. The first question was used to validate the amount of search time for self-generated reports against AI-generated reports. The results show one implicit connection with minimal outliers (Figure 1).

It is observed that when self and AI-generated search times are both high or low, the corresponding need for additional search is lesser. It implies that the higher the difference is between two levels of search, the higher is the potential utility. The response from the second question was then plotted over self-generated search time and AI-generated search time in Figure 2(a). We remove the ‘high-high’ and ‘low-low’ cases and represent the difference of the two times of the rest separately in Figure 2(b).

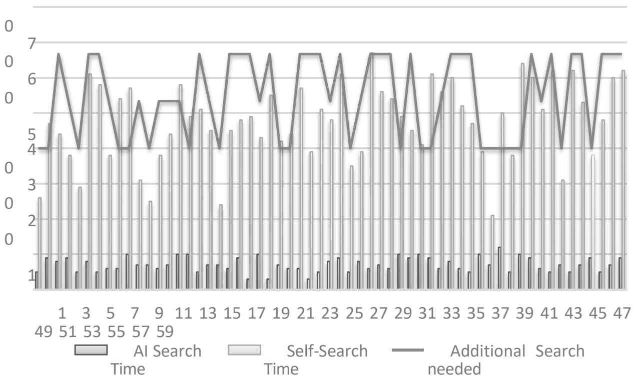

Fig. 1. Search Utility

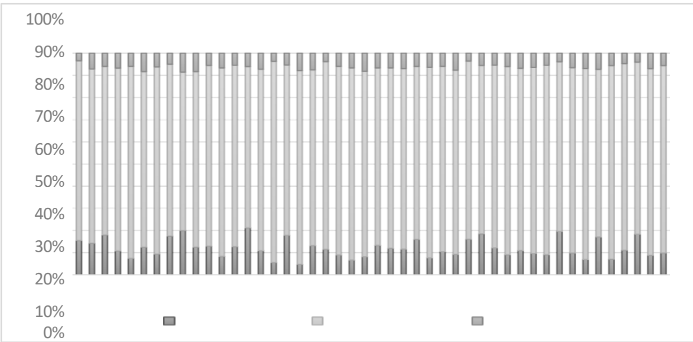

Fig. 2(a). Combined Impact

Fig. 2(b). Segregated Impact

We clearly observe a strong positive correlation between the additional search and the perceived impact. It signifies higher utility from higher search. But the theory is limited by the structure and profile of responses, and even the type of report generated is too narrow and homogeneous. The experiment must be extended to a more heterogeneous group with lesser to higher complexities and varied types in the reports. From such an experiment, we can understand the marginal rate of substitution (MRS) between self-search and AI search. In such a scenario, it would be imperative to give the participants complete freedom to combine natural and AI searches in various reports. The proportion of each search and the corresponding utility would lead to the final indifference curve.

4 Conclusion

The apparent motivation of the experiment was to work on a hypothesis that the level of self- satisfaction derived from AI-generated content cannot exceed that of self-generated content. If the hypothesis stands true, the implicit cost associated with generative AI would be (in effect) exaggerated; the user would gain lesser utility from the product as a whole. As a result, the market price at which the AI product is (and will be) traded, will be closer to the cost price and therefore, become highly volatile.

The experiment categorically indicated that the hypothesis is true to a great extent. Even those who were experts in using generative AI, showed much less ownership, pride and enrichment in AI-generated contents. Only in the incentivised group, it was observed that the enrichment level was comparatively higher. But this finding was diluted by the fact that the participants readily infringed with the generated content by re-generating them or by adding their own input. It is quite obvious that if AI is generating a content, ownership and pride will not be significant. But if the participant (provoked by some incentive) combines his/her own ability with the AI content, he/she should feel highly attached with the creation. The current experiment, however, did not provide complete clarity. It was observed that 'specific-domain' topics create higher ownership and pride, when compared with 'generic' topics simply because, 'domain' topics were complex in nature, and the generation process required higher expertise. Therefore, the satisfaction generated therefrom could possibly had been of the *expertise* rather than had been the *content*. There was little scope of observing this parameter. Therefore, based on the panel, and limited by the scope, we can apparently conclude that AI-generated topics contain lesser ownership and pride.

It is important to note that the panels were asked only to generate a small report of around two paragraphs. When this task complexity rises by a large extent; perhaps to generating a detailed article, or even a project report; the amount of natural search required might offset the utility derived from the content. As of now, it is logical to say that expert users are likely to be more satisfied with artificial content, if it only takes care of the 'search'. To validate this completely, the paper proposed to find out the MRS between natural and artificial search. But this requires larger heterogeneity in user demographics, and also in the complexity and volume of the topic. At the user level, clusters must be made based on technical expertise, technology adoption level, time availability, work profile and the like. At the content level, focus should be on creativity, volume, urgency, privacy and information availability. Bringing all these facets under one umbrella, we can expect to precisely generate the indifference curve, and subsequently the indifference map.

End notes:

1. https://explodingtopics.com/blog/chatgpt-users (accessed on 31 December 2023)

2. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-
02-01 (accessed on 31 December 2023)

3. See Davis (1989), Venkatesh et al. (2003) and Gursoy (2019).

References

Bai, L., Liu, X., & Su, J. (2023). ChatGPT: The cognitive effects on learning and memory. *Brain-X, 1*(3), e30. https://doi.org/10.1002/brx2.30

Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM, 64*(7), 58–65. https://doi.org/10.1145/3448250

Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society*, *75*, 102370. https://doi.org/10.1016/j.techsoc.2023.102370

Brown, G., Pierce, J. L., & Crossley, C. (2014). Toward an understanding of the development of ownership feelings. *Journal of Organizational Behavior*, *35*(3), 318–338. https://doi.org/10.1002/job.1869

Choi, J. H., Hickman, K. E., Monahan, A., & Schwarcz, D. (2023). ChatGPT goes to law school. *Journal of Legal Education*, *71*, 387. http://dx.doi.org/10.2139/ssrn.4335905

Davis, F. D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly*, *13*(3), 319–340. doi: 10.2307/249008

Drolet, A., Simonson, I., & Tversky, A. (2000). Indifference curves that travel with the choice set. *Marketing letters*, *11*, 199–209. https://doi.org/10.1023/A:1008198925117

Fernandes, A. S. (2012). Writing to learn writing skills: A case study. *European Journal of Engineering Education*, *37*(2), 179–192. https://doi.org/10.1080/03043797.2012.672964

George, A. S., George, A. H., & Martin, A. G. (2023). ChatGPT and the future of work: A comprehensive analysis of AI's impact on jobs and employment. *Partners Universal International Innovation Journal, 1*(3), 154–186. https://doi.org/10.5281/zenodo.8076921

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608*.
https://doi.org/10.48550/arXiv.1812.04608

Huurdeman, H. C., Wilson, M. L., & Kamps, J. (2016, March). Active and passive utility of search interface features in different information seeking task stages. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval* (pp. 3–12). https://doi.org/10.1145/2854946.2854957

Kaur, J., & Saini, J. R. (2014). Emotion detection and sentiment analysis in text corpus: A differential study with informal and formal writing styles. *International Journal of Computer Application, 101(9),* 1–9.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187-192. http://dx.doi.org/10.2139/ssrn.4375283

Rice, S., Crouse, S. R., Winter, S. R., & Rice, C. (2024). The advantages and limitations of using ChatGPT to enhance technological research. *Technology in Society*, *76*, 102426. https://doi.org/10.1016/j.techsoc.2023.102426

Schwabe, H., & Castellacci, F. (2020). Automation, workers' skills and job satisfaction. *PloS one*, *15*(11), e0242929. https://doi.org/10.1371/journal.pone.0242929 Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT–Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, 74. https://doi.org/10.1016/j.ijinfomgt.2023.102700

Venkatesh, V., Morris, Michael G., Davis, Gordon, B., Davis, Fred D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3): 425–478. https://doi:10.2307/30036540

Wang, X., Lin, X., & Shao, B. (2022). How does artificial intelligence create business agility? Evidence from chatbots. *International journal of information management*, *66*, 102535. https://doi.org/10.1016/j.ijinfomgt.2022.102535

Yilmaz, E., Shokouhi, M., Craswell, N., & Robertson, S. (2010). Expected browsing utility for web search evaluation. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (pp. 1561–1564). https://doi.org/10.1145/1871437.1871672