The Differences in the Use of Passive Voice in Academic and Popular Writing on Artificial Intelligence and Machine Learning

Aruzhan Tolegen

School of Science and Humanities, Nazarbayev University

WCS 230: Say What You Mean: Clarity, Precision, and Style in Academic Writing

Jeffrey Carney

1 December 2023

Abstract

This research investigates the nuanced use of passive voice in academic and popular writing on Artificial Intelligence (AI) and Machine Learning (ML). Through a comprehensive analysis of articles from diverse platforms, encompassing reputable journals and popular sources, the study explores patterns in passive voice usage, considering both quantitative frequency and qualitative functions. Drawing on the insights of scholars such as Bacon, Wang, Hiltunen, Hyland, Jiang, Ding, and Williams & Bizup, the research has distinct rhetorical functions of passive voice, including focusing on experiments, impersonalization, and objectivity. Also, the study reveals a divergence in the frequency of passive voice between academic and popular writings. Academic papers have a higher frequency of passive voice, aiming to emphasize experimental outcomes and achieve objectivity. In contrast, popular writings employ passive voices to underscore factual information to a broader audience.

The Differences in the Use of Passive Voice in Academic and Popular Writing on Artificial Intelligence and Machine Learning

The exploration of passive voice usage in written communication has been a subject of enduring interest, particularly in academic and popular writing. As Artificial Intelligence (AI) and Machine Learning (ML) continue to shape contemporary discourse, understanding the diverse rhetorical functions of passive voice and their frequency choice in these domains becomes essential. This research attempts to identify patterns – even subtle trends or nuances – that illuminate distinct roles played by passive voice and how frequently authors prefer to use the passive voice in their writings across academic and popular writings on AI and ML.

Many authors prefer to use an active voice in sentences because of the natural subject-verb-object structured language. However, the passive voice can be a better option and preform some functions more effectively. Academic investigations into passive voice usage provide a foundational understanding of its roles in scholarly discourse. Bacon (2013) briefly stated the three fundamental functions of the passive voice: to conceal the doer or performer, to employ the subject position to identify a thing or idea that has been introduced in a previous sentence and "emphasize the absence of agency and power in a person or thing that is acted upon" (p. 45). However, they are not the only functions of the passive voice. Wang (2009), for instance, explored two academic corpora, uncovering at least ten distinct rhetorical functions in academic writing. These functions, as identified by Wang, exhibited interconnectedness, portraying the nature of the passive voice's deployment in scholarly communication.

Hiltunen's (2016) comprehensive analysis of a large English for Academic Purposes (EAP) corpus contributes further insights. The study, spanning diverse disciplines, discerns a significant variance in passive constructions between hard and soft sciences. This observation,

substantiated through various frequency measurement methods, not only reaffirms prior qualitative findings but also provides empirical evidence for the prevalence of passive voice in different academic domains. **The analysis of passive voice patterns across disciplines, a comprehensive summation of linguistic choices, provides crucial information that will significantly help understand the differences in using passives within an academic context, since research about AI/ML is often associated with hard sciences [2].**

Hyland and Jiang (2017) used a corpus of 2.2 million words from leading journals across four disciplines, the research focuses on the evolution of informal elements in academic writing, particularly in the context of interpersonal interactions. Since these findings reveal an increase in informality, driven mainly by using first-person pronouns, unattended references, and sentences starting with conjunctions, I want to investigate how informality manifests through passive constructions in my corpora.

Ding's work, spanning both 2002 and 1998, focused on the social values shaping scientific writing. Ding (2002) argues that passives in scientific communication are influenced by the values of falsifiability and cooperation among scientists. The need for repeatability and cooperation, as asserted by Ding, shapes the prevalence of passive voice in scientific discourse. Ding's earlier work in 1998 emphasizes the impact of "rationality and reason" on the dominance of passive voice in scientific writing before the end of the 19th century, shedding light on the historical evolution of this linguistic construct within the scientific community.

Williams and Bizup (2014) enhance the discourse by identifying cases where authors, in an attempt to achieve objectivity through passive voice, fail to eliminate the presence of vivid characters and increase objectivity [3a].

This study's research question is aimed to find out the distinct patterns of passive voice usage in academic and popular writing in the context of Artificial Intelligence and Machine Learning [3b]. By analyzing the patterns in functions of passive voice usage, this paper aims to distinguish how the roles of the passive voice vary across academic and popular writings, providing a comprehensive understanding of its rhetorical functions in the discourse on AI and ML and contributing to the broader conversation on language use in the discourse surrounding AI and ML.

Method

Corpus

To maintain a comprehensive scope, I chose articles from various publication platforms, each offering a distinct perspective. Academic rigor is represented through articles from reputable journals such as *Nature Medicine*, *Complex & Intelligent Systems*, and the *Association for Computing Machinery*.

For the popular article, I chose from platforms like *ZDNET*, *Harvard Business Review*, and *Simplilearn.com*. By including articles from these diverse platforms, the corpus will represent a general category of articles. *ZDNET* may focus on the latest technological advancements, *HBR* on the business implications, and *Simplilearn.com* on educational content. This variety ensures that the analysis of passive voice in popular writing is not limited to a specific niche but shows a more extensive scale representation of how AI and ML topics are conveyed to the general audience.

The selected timeframe spans recent years, with publications ranging from 2020 to 2023, reflecting recent developments in the field and showing the relevance of the information. For a detailed breakdown of the corpus, refer to the Appendix.

Procedure

For the data analysis, I will use the embedded tools within Google Chrome to locate and identify passive clauses in the research and popular articles. Specifically, I will employ the browser's search functionality to identify keywords indicative of passive constructions, such as "was," "were," "to be," and "being." + past particle construction. All six articles will be analyzed according to this procedure. Besides that, using the same technique, I will count the number of first-person pronounce used by the authors. However, only part of the article will be considered. I will exclude the abstracts, tables, figures, and direct quotes from other research that authors referred to in their writings. This approach avoids reading every sentence, simplifies the process, and reduces the likelihood of overlooking instances of passive voice. Then, I will identify the functions of the passive voices used in the articles. After that, I will compare the results obtained through qualitative and quantitative analysis.

Results

Consider these two sentences from the academic papers where the passive voice is employed:

1(a). Gradient descent **was used** to train a model embedded in the system (Spring et al., 2020).

1(b). This study mainly focuses on different AI/ML techniques that **were applied** for the COVID-19 outbreak (Doğan et al., 2021).

In all sentences, the type of passive voice is finite passive. Because the verbs “was used” and “were applied” are in the past participle form. By not explicitly stating the agent, the focus remains on the outcome, aligning with the project's goal of emphasizing the experimental results rather than the specific actors involved (Hiltunen, 2016, p. 136). And this is unimportant because, in those examples, we clearly understand that the performers are the researchers themselves. This clear example is the function of “focusing on the experiment,” which was described by Wang (2009).

The same logic can be applied to popular writings; while explaining some concepts, authors may avoid mentioning the performer to “focus on the action” or “focus on the result”.

2(a). ML can target a marketing campaign to customers who are more likely to respond, or disallow credit card transactions that are likely **to be fraudulent** (Siegel, 2023).

2(b). Finally, DeepMind, a pioneer in the fields of AI and deep learning, **was established** in September 2010 (Terra, 2023).

Bacon stated that hiding a performer is the fundamental function of the passive voice. This is true, but the reason why exactly authors are hiding the actor is to make the sentence more impersonal Wang (2009). Consider examples 1 and 2, “will be fixed” and “be developed.” we know that the researchers must do these actions, but by avoiding showing the actor, the writer tries to avoid discussing and emphasizing their relation to it.

3(a). The vulnerability research is internal to product development, and **will be fixed** without being made public (Doğan et al., 2021).

3(b). A neural network can also **be developed** to extract the visual features of this disease (Vaishya et al., 2020).

In popular writing, the passive voice does not aim to hide and completely avoids mentioning the performer when the performer is the author/author of the article. Even while using the passive voice, the performer can be seen in the sentence, like in the example below. In this case, the goal is to emphasize the actions (reading and editing) that can be performed on the knowledge base rather than explicitly state who performs these actions.

4(a). It's a free and open knowledge base that can **be read and edited** by both humans and machines (Anadiotis, 2020).

Usually, numerical, graphical information obtained during the research is represented in graphs and tables. This function is presented in all academic articles that were analyzed. Since "Passive voice also tells the reader where to find related information," most authors of academic writings use passive voice in this function, see examples 5(a) and 5(b) (Wang, 2009). No examples of this function were found in popular writings.

5(a). The steps taken in this study **are shown in** *Fig. 1.* (Doğan et al., 2021).

5(b). The ML system includes all of the components **illustrated in** *Figure 1.* (Spring et al., 2020)

To make the paper more objective, authors try not to mention their opinion explicitly, so they delete the first person's pronouns and use the passive voice (Wang, 2009). You can see it in example 6(a); the passive form makes the sound more objective. Its active equivalent will focus on the author's subject impression: "We hope that AI will reduce medical costs."

6(a). **It is hoped** that AI will reduce medical costs **(**Rajpurkar et al., 2022).

Table 1

The Number of Passive Clauses per 1000 Words

Source	**# of passive clauses**	**# of passive clauses per 1000 words**

Academic	Doğan et al. (2021)	139	18.5
	Spring et al. (2020)	97	10.9
	Vaishya et al. (2020)	15	12.5
	Rajpurkar et al. (2022)	28	11.6
Popular	Anadiotis (2020)	16	6.4
	Siegel (2023)	5	7.27
	Terra (2023)	16	11.1

In academic papers, such as Doğan et al. (2021), Spring et al. (2020), Vaishya et al. (2020), and Rajpurkar et al. (2022), the number of usage of passive clauses demonstrates a range from 15 to 139 instances. **Considering the frequency, or the occurrence of passive voice, within a given set of 1000 words, the ratio varies from 10.9 to 18.5 passive clauses [1]**. On the other hand, the analysis of popular writings, including Anadiotis (2020), Siegel (2023), and Terra (2023), portrays a generally lower number ranging from 5 to 16 instances, with a corresponding ratio of 6.4 to 11.1 passive clauses per 1000 words. This information explains the authors' linguistic preferences and the structural choices made in scholarly discourse.

Table 2

The Number of First Person pronounces

Source		**# first person pronounces**	**# first person pronounces per 1000 words**
Academic	Doğan et al. (2021)	0	0
	Spring et al. (2020)	40	4.5
	Vaishya et al. (2020)	2	1.6
	Rajpurkar et al. (2022)	19	7.8
Po	Anadiotis (2020)	44	17.6

pul ar	Siegel (2023)	7	10.2
	Terra (2023)	3	2.8

In academic papers the presence of first-person pronouns normalized to account for article length, the rate per 1000 words varies from 1.6 to 7.8. Conversely, in analyzing popular writings, the frequency of first-person pronoun usage in popular articles ranges from 2.8 to 17.6 first-person pronouns. This difference in the use of first-person pronouns between academic and popular genres suggests a greater inclination toward a personal and engaging writing style in popular articles.

Discussion

The results of our study, reveal significant differences in the use of passive voice between academic and popular writing on the subject of Artificial Intelligence and Machine Learning.

This discrepancy in the results suggests that each genre's distinct communicative goals and audience expectations influence linguistic choices. Hyland (2017) stated that academic papers are becoming more informal to “involve readers in arguments and secure support for their claims through positioning themselves more explicitly concerning their ideas and readers” (p. 49). However, as we can see from the obtained results, academic papers about AI/ML do not conform to this evolving trend. Most of the passive voice used in the academic paper about AI/ML performs the following functions: “focusing on the experiment,” “Impersonalization,” and “Objectivity: Focusing on the experiment.”. Contrary to the emergent informality, most authors in academic papers prefer the passive voice to position themselves less. The choice to use passive voice in academic papers shows the intention to minimize the sense of authorial

presence, a goal not adhered to by authors of popular writings, leading to a lower frequency of passive voice in their discourse.

The frequency of passive clauses per 1000 words in academic papers ranges from 10.9 to 18.5, while in popular writing, it falls within the range of 7.27 to 11.1. Similarly, the usage of first-person pronouns per 1000 words varies between academic and popular articles, ranging from 1.6 to 7.8 in academic papers and 2.8 to 17.6 in popular writing. From this we see the following trend in academic and popular articles: authors with a relatively high frequency of passive voice have a low frequency of first-person person pronounciation.

Not mentioning the agent helps achieve the sentence's object-centered sense (Ding, 2002). This says that authors of academic papers try to increase passive voice and delete first-person pronouns to be more objective. An example of this can be seen in 6(a). However, Williams & Bizup (2014, p. 59) stated that deleting an "I or we don't make a researcher's thinking more objective." Objectivity is not about just hiding the actor**,** but as stated by Ding (1988, p. 118), "It represents the world in terms of objects, things, and materials. Thus, it appears to be more 'object-oriented' and 'thing-oriented' than subject-oriented and human-oriented". So objectivity is presenting information in a manner that reflects the external world independently of personal biases or perspectives.

In contrast, academic papers do replace the first person pronoun from their articles consequently, they do not use passive voice to make their paper sound more objective. Most of the time, they emphasize essential parts of the sentence.

The next difference in the use of passive voice becomes particularly evident when considering its application in conveying results. In academic writing, authors frequently deploy

passive voice constructions to emphasize the outcomes of experiments, as in examples 2(a), and 2(b). This practice aims to direct the reader's attention to the obtained results.

Conversely, in popular writing, the use of passive voice serves a distinct communicative function. Rather than emphasizing the process or method of obtaining results, authors in popular writings especially use passive voice constructions to highlight the factual nature of the presented information, emphasizing communicating the obtained results to a broader audience. This distinction is evident when examining specific examples, such as those presented in examples 1(a) and 1(b).

Also, in academic papers, the passive voice sometimes serves the function of "Indicating figures, tables, or other related information" (Wang, 2009). On the contrary, authors of popular writings practically never use passive voice in this manner.

This research finds that this difference in the application of passive voice is not uniform across genres but is instead shaped by the specific communicative goals and audience expectations inherent in academic and popular discourses. Eventually, this research will fill in the gap of passive voice usage in popular writings since there is little research conducted about them. [2582 words].

References

Bacon, N. (2013). *The well-crafted sentence: A writer's guide to style* (2nd ed.). Bedford, 36–56.

Ding, D. D. (2002). The passive voice and social values in science. *Journal of Technical Writing and Communication*, *32*(2), 137–154. https://doi.org/10.2190/efmr-bjf3-ce41-84kk

Ding, D. D. (1998). Rationality reborn: Historical roots of the passive voice in scientific discourse. In J. Battalio (Eds.), *Essays in the Study of Scientific Discourse: Methods, Practice, and Pedagogy* (pp. 117–135). Ablex Publishing Company.

Hiltunen, T. (2016). Passives in Academic Writing: Comparing Research Articles and Student Essays Across Four Disciplines. In López-Couso, M. J., Méndez-Naya, B., Núñez-Pertejo P. & Palacios-Martínez M. I. (Eds.) *Corpus linguistics on the move: Exploring and understanding English* (pp. 132–157). Brill. https://doi.org/10.1163/9789004321342_008

Hyland, K., & Jiang, F. (2017). Is academic writing becoming more informal? *English for Specific Purposes*, *45*, 40–51. https://doi.org/10.1016/j.esp.2016.09.001

Wang, H. (2009). Passive voice in research articles: Revisiting rhetorical functions. *Proceedings of the 26th International Conference on English Teaching and Learning in the R. O. C,* 774-784.

Williams, J. M., & Bizup, J. (2014). *Style: Lessons in clarity and grace* (11th ed.). W&B.

Appendix

List of Sources Under Investigation

Anadiotis, G. (2020, December 4). *Getting there: Structured data, semantics, robotics, and the future of AI.* ZDNET. https://www.zdnet.com/article/what-is-machine-learning-everything-you-need-to-know/

Doğan, O., Tiwari, S., Jabbar, M. A., & Guggari, S. (2021). A systematic review on AI/ML approaches against COVID-19 outbreak. *Complex & Intelligent Systems*, *7*(5), 2655–2678. https://doi.org/10.1007/s40747-021-00424-8

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

Siegel, E. (2023, March 24). How machine learning can improve the customer experience. *Harvard Business Review.* https://hbr.org/2023/03/how-machine-learning-can-improve-the-customer-experience?aut ocomplete=true

Spring, J. M., Galyardt, A., Householder, A. D., & VanHoudnos, N. (2020). On managing vulnerabilities in AI/ML systems. *Association for Computing Machinery*. https://doi.org/10.1145/3442167.3442177

Terra, J. (2023, May 30). *10 years of artificial intelligence and machine learning*. Simplilearn.com. https://www.simplilearn.com/ten-years-of-artificial-intelligence-and-machine-learning-art icle

Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for covid-19 pandemic. *Diabetes, Metabolic Syndrome: Clinical Research, Reviews*, *14*(4). https://doi.org/10.1016/j.dsx.2020.04.012

[1] Appositive that performs any function besides identifying a person

An appositive is a noun phrase that provides extra details about the term (Bacon, 2013, p. 125). In the example from the paper, the appositive defines a term "the frequency, or the occurrence of passive voice, within a given set of 1000 words". As Bacon (2013, p. 126) mentioned, technical term definitions can appear in an appositive. In the example, the term "frequency" definition appears in an appositive.

[2] Summative modifier effectively

According to Williams & Bizup (2014, p. 154) the summative modifier construction allows the writer to bring a sense of closure to the initial idea and then continue with more specific details in the relative clause. Which is seen in the sentence 2. In sentence [2] the phrase "a comprehensive summation of linguistic choices" serves as a summative modifier, summarizing the nature of the analysis.

[3] Shaped effectively

The sentence [3a] with 30 words is shaped effectively. According to Williams & Bizup (2014, p. 145), for a sentence to be effectively shaped, it should first get to the subject quickly. Second, Get to the verb and object quickly. This sentence gets to the subject quickly ("Williams and Bizup"), clarifying who is acting. Similar to the sentence [3b], where the subject is effectively introduced at the beginning of the sentence ("this study's research question"), followed by a clear verb ("aimed to find out"). The sentence structure is clear and direct, making conveying the main idea practical.

Reference list

Bacon, N. (2013). *The well-crafted sentence: A writer's guide to style* (2nd ed.). Bedford, 36–56.

Williams, J. M., & Bizup, J. (2014). *Style: Lessons in clarity and grace* (11th ed.). Pearson.