Artificial Intelligence, Social Computing and Wearable Technologies, Vol. 113, 2023, 216–225
https://doi.org/10.54941/ahfe1004194

To What Extent Can AI Simplify Academic Paper Writing?

Youji Kohda and Amna Javed

Japan Advanced Institute of Science and Technology, Nomi, Ishikawa 923-1292, Japan

ABSTRACT

In the AI era, how will the work of researchers change and what will the future of the research profession be like? This study discusses how academic paper writing can be made more efficient through AI assistance. To facilitate our discussion, we used a somewhat unusual approach. Noting that the novelty and inventive step of an invention and the originality and academic significance of an academic paper are the same, we propose a six-step model for AI-assisted academic paper writing, inspired by AI-assisted patent prior art search. On the basis of the proposed model, we present our thoughts on the future of researchers' work and the research profession in general.

Keywords: Artificial intelligence, Professional work, Future of work, Researchers, Academic paper writing, Prior literature review, Patent prior art search

INTRODUCTION

It has been pointed out that the advent of AI will result in the loss of jobs, in particular, white-collar jobs (e.g., Frey and Osborne (2017)). What about the research profession? There is no rule without exception, but is the research profession *the* exception? Are they secure because they are intellectual occupations?

A researcher's achievement is a published article, preferably in a peer-reviewed academic journal. Academic papers are written for experts with prior knowledge, but even experts cannot be experts in everything, so they must be written in a way that anyone can read and understand them to pass peer review. This is what makes them different from novels which tell an imagined story. To be understandable, the descriptions of academic papers should be written in accordance with a format many people will find easy to understand, i.e., IMRaD (Introduction, Materials and methods, Results, and Discussion) (e.g., Nair and Nair (2014)).

The introduction of IMRaD typically consists of an introductory section describing the background and objectives of the research and a prior-literature-review section. A research question (RQ) is sometimes positioned at the end of the introduction. An RQ is a brief question that presents the purpose of the study: what issue is being clarified in the research. The materials and methods of IMRaD describes what the researcher did to answer the RQ. While an RQ becomes more valuable when it is unique, materials and methods are not necessarily unique because it may be more

appropriate to use the same materials or methods as in previous studies to make a study reproducible. The discussion of IMRaD examines the results obtained to check to what extent the RQ in the introduction has been answered. In most cases, the RQ will never be completely answered, and the remaining unanswered portion will be mentioned in the conclusion as future work. Note that the structure of the paper is nested: the outer pair of introduction and discussion describe the presentation of an RQ and the responses to the RQ. The inner pair of materials and methods and results describe what the researcher did and the results. Thus, IMRaD is a good guide to understanding the content of a paper.

The focus has been on AI that specializes in natural language processing, triggered by ChatGPT (OpenAI, 2022). If one wants to understand the content of a document, one generally has to read it. However, with AI such as ChatGPT, not only can a rather long document be summarized into shorter sentences, but it is now possible to understand an overview of the document interactively by asking questions, as if one were asking questions of a human who understands the document.

We would like to point out here that it is no longer difficult for AI to divide and summarize an academic paper into introduction, materials and methods, results, and discussion, since academic papers are written with the IMRaD format. This suggests that while it may be difficult for AI to decipher why a novel is interesting, it is easy for it to decipher the gist of an academic paper. Thus, the emergence of AI specialized in natural language processing will ultimately have a large impact on the job of research.

Writing academic papers is an important task for researchers to build up their achievements. Writing an academic paper is time consuming, especially its prior literature review. For an academic paper to be the accepted for peer review, the researcher must convince reviewers in the same field that the content of the paper is original and academically significant. The most direct way to claim originality and academic significance is to show that the content of the paper has not been published in any previously published paper (i.e., it is original) and that it cannot be derived from a simple combination of previously published papers (i.e., it is academically significant). To properly confirm the originality and academic significance, it is necessary to examine previously published papers as comprehensively as possible. Therefore, the emergence of AI specialized for natural language processing is expected to be a means of streamlining this prior literature review.

However, relying on AI to conduct prior literature review may hinder the ability of researchers to read papers, and may even result in the loss of their ability to write papers (see also Kohda (2020, 2022)).

On the basis of the above discussion, the RQs of this study are as follows:

RQ1 How does an AI specialized in natural language processing streamline the process of writing academic papers? Specifically, how will it streamline the process of prior literature review?

RQ2 When prior literature research is made more efficient through AI, will the researcher's ability to write academic papers decline or, conversely, will it be enhanced?

To answer RQ1, we propose a six-step model for AI-assisted academic paper writing, drawing inspiration from the process of AI-assisted patent prior art search and argue that writing academic papers can be made more efficient through AI. To answer RQ2, we argue that the ability of researchers to write papers can be enhanced with the help of AI that can find "hit papers" that match their interests.

LITERATURE REVIEW: OVERVIEW

A literature review is the task of analyzing a set of previously published literature and summarizing what we know and do not know thus far about the RQ. We offer the following three topics on literature review that are relevant to the discussion of this study.

(1) According to Snyder (2019), prior literature review can be categorized into three typologies: systematic, semi-systematic, and integrative. Which typology should be used depends on the nature of the RQ. A systematic review synthesizes the results of prior literature with the same RQ to obtain a definitive answer to the RQ. Meta-analysis is a statistical technique that integrates the results of multiple statistical analyses to determine whether a factor or factors are related to a particular result. A semi-systematic review classifies what research has been done and how research has evolved to answer the RQ. A method similar to qualitative analysis is used for classifying prior literature. When there are very few previous papers that attempted to answer the RQ, an integrative literature review attempts to find a path to answer the RQ in a heuristic manner by synthesizing the findings of previous prominent papers. The path to answering the RQ can be presented in the form of taxonomies, theoretical frameworks (conceptual frameworks), etc.

These three typologies of prior literature review are continuous and not clearly distinguishable. Systematic and integrative are the two extremes, with systematic being deductive and integrative being inductive. The prior literature review of most academic papers is not so extreme and can be classified as semi-systematic. Therefore, the remainder of this study focuses on cases in which the prior literature review is semi-systematic.

(2) In a semi-systematic review, the prior-literature-review section should avoid paper-by-paper explanations and adopt topic-by-topic explanations. For example, after analyzing 50 papers, if 4 topics are identified, each topic should be described in turn. The researcher should explain each topic in their own words and cite the related prior literature as evidence. In other words, a literature review is a task to read a set of papers, paper-by-paper, and rewrite them topic-by-topic. We call the paper-by-paper reading stage literature analysis and the topic-by-topic rewriting stage literature synthesis.

(3) A typical question of a beginner is "How many papers should I read for a prior literature review?" According to Shon (2015), one should read 30 to 50 papers for a prior literature review. In the beginning, he says, you will discover something new each time you read them, but as you read more, you will start to feel that you have already read the

content somewhere else, and some classifications will naturally come to your mind. Toward the end, he says, you will not encounter any new content, but you should read through to the end for confirmation. The classifications that come to your mind during the readings are the topics, and by explaining the topics in order, you can write topic-by-topic explanations.

The above Shon's answer to the typical beginner's question suggests that literature analysis and literature synthesis are conducted simultaneously in the mind when a human researcher conducts a prior literature review. Could some of the work be delegated to AI to lower the burden of the prior literature review? What about leaving the task of literature analysis to AI and enabling human researchers to concentrate on the task of literature synthesis? Because AI can process documents without limitation, it also creates the advantage of reducing the risk that some topics might be overlooked when limited to 50 papers due to the constraints of human reading ability.

METHOD: LEARNING FROM PATENT PRIOR ART SEARCH

To facilitate our discussion, we used a somewhat unusual approach. We argue that the nature of a patent prior art search is analogous to a prior literature search in academic research. Combining this point with the fact that services for AI-assisted patent prior art search already exist, we conclude that AI-assisted prior literature searches can be developed.

In a patent prior art search, an inventor first examines whether their idea (technological innovations) is patentable. There are two conditions for patentability: novelty and the inventive step, which must be satisfied simultaneously. Novelty means that there is no patent document in the patent office database that encompasses the inventor's idea (Figure 1a). The inventive step is that the inventor's idea cannot be reproduced by a simple combination of patent documents registered in the patent office database, i.e., a simple combination of patent documents cannot cover the inventor's idea (Figure 1b).

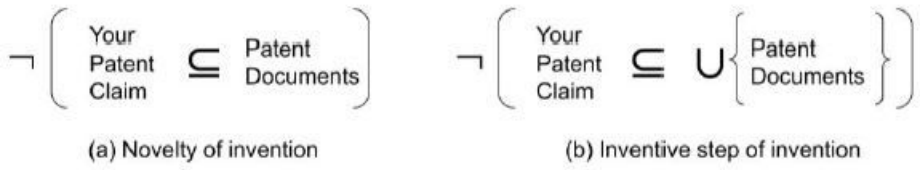

Figure 1: Patentability of invention over prior art.

In the introduction, we stated that academic papers need to have originality and academic significance. In fact, the novelty of an invention and originality of an academic paper and the inventive step of an invention and academic significance of an academic paper are essentially the same (Figure 2). In a patent, the claims are searched in the patent office database, while in an academic

paper, the IMRaD is searched in the abstracts and citations database, which is a collection of academic journals.

Figure 2: Inventive step of IMRaD over prior literature.

Determining the novelty of an invention or originality of an academic paper is not difficult using a computer. It is sufficient to check the patent office database or abstracts and citations database to make sure that there are no documents that describe the same idea or same IMRaD. To examine the inventive step of an invention or academic significance of an academic paper, however, one must consider combinatorial explosion. Combinatorial explosion occurs because we must consider the combination of patent documents or academic papers.

An AI service is commercially available which can heuristically determine the inventive step of an inventor's idea (AI Samurai, 2019; Kohda and Shirasaka, 2021). This service accepts an idea in natural language and searches the Japan, U.S., or China patent office database. The results are output in the form of a table that mimics a claim chart. Claim charts are used by patent attorneys to determine patentability. If even one simple combination of patent documents in the patent office database can be found that covers the input idea, then the idea is judged as non-inventive. If any simple combination of patent documents in the patent office database is unlikely to cover the input idea, then the idea is judged as inventive.

AI Samurai (2019) determines patentability on the basis of this claim chart and shows a judgment score: "A" being the most patentable, "B" and "C" being less patentable, and "D" being the least patentable. Figure 3 shows an inventive-step analysis result of invention US7136853 (Kohda and Endo, 2006) by AI (AI Samurai, 2019). This patent was filed in Japan on Sep. 7, 1995, and in the U.S. on Sep. 6, 1996, and was registered as a U.S. patent on Nov. 14, 2006. Figure 3a shows the result of the patentability determination in the U.S. patent database as of Sep. 7, 1995, and Figure 3b shows the result of the patentability determination in the U.S. patent database as of Sep. 7, 1997. These results indicate that the invention had an inventive step and was patentable as of Sep. 7, 1995, because the judgment score was "B." However, it lost its inventiveness with the passage of time. In fact, as of Sep. 7, 1997, the judgment score changed to "D."

We present our proposed six-step model for AI-assisted academic paper writing in the next section. We believe this model can be implemented as a practical service. Since AI Samurai (2019) can determine novelty and the inventive step in patent prior art searches, an AI should be able to determine originality and academic significance in academic papers through prior literature searches.

Judgment result: B (patentable) as of September 7, 1995

US7136853 Claim 1	US5444842	US5423037	US5444840	US428778	US5428773
An information providing method ..., comprising;	70%	72%	49%	36%	32%
receiving retrieval information for a user ...;	27%	38%	55%	34%	32%
selecting one piece of advertising information ...;	28%	18%	36%	26%	17%
and transmitting the selected advertising ..., wherein	50%	58%	36%	35%	34%
said retrieval information includes one of	35%	24%	32%	38%	24%

Figure 3a: Inventive-step analysis of US7136853 (Kohda and Endo, 2006) through AI (AI Samurai, 2019) as of Sep. 7, 1995.

Judgment result: D (no patentability) as of September 7, 1997

US7136853 Claim 1	US5659746	US5649742	US5664110	US5560005	US5657450
An information providing method ..., comprising;	78%	81%	100%	76%	81%
receiving retrieval information for a user ...;	73%	83%	38%	66%	53%
selecting one piece of advertising information ...;	40%	46%	45%	55%	49%
and transmitting the selected advertising ..., wherein	61%	57%	96%	45%	58%
said retrieval information includes one of	39%	45%	37%	30%	46%

Figure 3b: Inventive-step analysis of US7136853 (Kohda and Endo, 2006) through AI (AI Samurai, 2019) as of Sep. 7, 1997.

PROPOSAL: AI-ASSISTED ACADMIC PAPER WRTING MODEL

Figure 4 presents the proposed six-step model for AI-assisted academic paper writing. The model takes full advantage of the fact that academic papers can be broken down into IMRaD using AI.

Step 1 is the exploratory phase for the researcher: searching for papers that fit their interest. This step is not formally part of the prior literature review but is an essential step for writing an academic paper. This step is completed when one encounters a "hit paper" that helps conceive of an original RQ. Once such a hit paper is found, one can find more hit papers by using the RQs, keywords, and references described in the first paper. The output of this step is the RQ the researcher wants to answer, and it will be the input for the prior literature review in Step 2. Note that this RQ is tentative and can and should be revised in the remaining steps.

Step 1 can be significantly assisted through AI. There are many possibilities such as visualizing the abstracts and citations database through the researcher's navigation or recommending relevant papers from the abstracts and citations database. As of 2023, several AI services capable of assisting with step 1 have been introduced to the market.

Steps 2 through 5 is the process of creating the skeleton of a new academic paper. The skeleton of an academic paper is the IMRaD format, and Figure 4 shows how IMRaD progresses. Step 2 forms the Introduction, Step 3 forms

Materials and methods, Step 4 forms Results, and Step 5 forms Discussion. In Step 6, a draft of the academic paper is produced from IMRaD. Note that these steps are data-driven, specifically IMRaD-driven. Who and how the IMRaD is created is a secondary issue.

Step 2 is the AI-assisted prior-literature-review step. The task of literature analysis in Step 2 can be significantly assisted through AI (as explained in the previous section). The task of literature synthesis in Step 2 is a task for a human researcher by scrutinizing the results of AI's analysis to determine to what extent the RQ in Step 1 has been answered with the prior literature. As already mentioned, literature analysis examines prior literature paper-by-paper, while literature synthesis organizes the analysis result topic-by-topic. AI-supported literature analysis can find multiple prior papers covering the IMRaD of the hit paper(s) identified in Step 1 as well as in Figure 3 in a patent prior art search (see also Figure 5). Note that these prior papers provide hints for identifying topics. With the assistance of AI, a human researcher would be able to identify a potential topic by interactively summarizing a set of similar literature to a prior paper. This process can be repeated for the prior papers found in the literature analysis, and these topics become candidates for the topics in the literature synthesis. In this way, AI can also assist with literature synthesis that human researchers do.

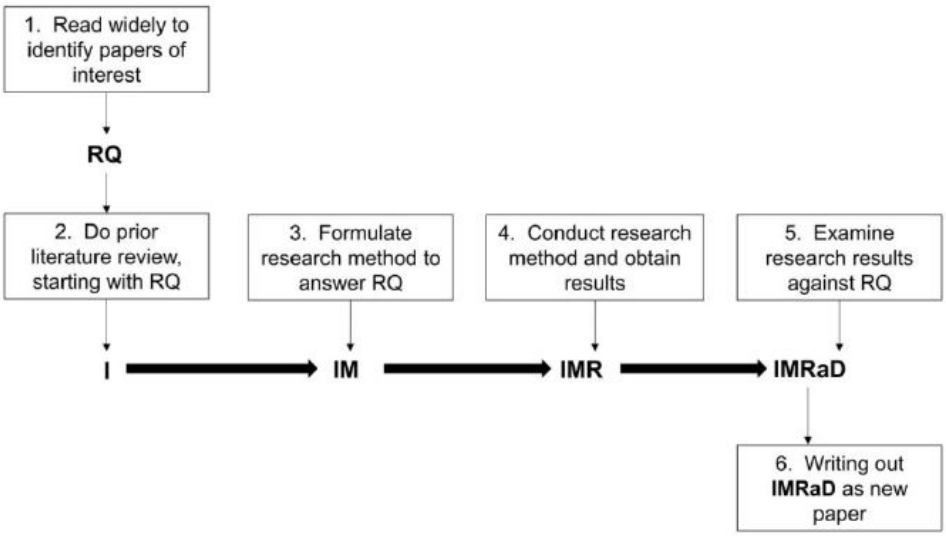

Figure 4: Six steps from conception to writing new paper.

Step 3 is formulating a research method to answer the unanswered portion of the RQ identified in Step 2. Step 4 describes the results obtained by conducting the research method. Step 5 examines the results of Step 4 to determine to what extent they answer the unanswered portion of the RQ. The size of unanswered portion in Step 2 is expected greatly decrease in Step 4. Steps 4 and 5 are the core of the research conducted by the researcher, so they are the steps in which the researcher's competence is exclusively tested. Once Step 5 is completed, the IMRaD of a new paper is ready for drafting in Step 6.

Let us elaborate on Step 6. Figure 5 corresponds to the claim chart in the patent search (Figure 3). In a patent prior art search, the claims of an invention are the target of analysis, while in Step 6, the IMRaD of the new paper is the target of analysis. Note that Figure 5 compares with two prior papers but could be compared to many more. The natural targets for the analysis of academic significance are the papers that were analyzed in Step 2; alternatively, the abstracts and citations database can be searched once again to refresh prior literature that will be cited in the new paper.

In Step 6, it is necessary to know how the prior literate relates to the new paper and cite the previous papers in appropriate sections in the new paper. Figure 5 shows the degree of overlap between the IMRaD of the new paper and prior literature. The numbers in the table show the possible overlap between all the IMRaD components of the two papers. Large numbers mean that there is large content overlap, and the prior literature must be cited where appropriate in the new paper. Table 1 shows a guideline of when and how the prior literature should be cited in the new paper.

IMRaD of a new paper	IMRaD of prior paper 1				IMRaD of prior paper 2			
	I	M	R	D	I	M	R	D
I	$n_{1II}\%$	$n_{1IM}\%$	$n_{1IR}\%$	$n_{1ID}\%$	$n_{2II}\%$	$n_{2IM}\%$	$n_{2IR}\%$	$n_{2ID}\%$
M	$n_{1MI}\%$	$n_{1MM}\%$	$n_{1MR}\%$	$n_{1MD}\%$	$n_{2MI}\%$	$n_{2MM}\%$	$n_{2MR}\%$	$n_{2MD}\%$
R	$n_{1RI}\%$	$n_{1RM}\%$	$n_{1RR}\%$	$n_{1RD}\%$	$n_{2RI}\%$	$n_{2RM}\%$	$n_{2RR}\%$	$n_{2RD}\%$
D	$n_{1DI}\%$	$n_{1DM}\%$	$n_{1DR}\%$	$n_{1DD}\%$	$n_{2DI}\%$	$n_{2DM}\%$	$n_{2DR}\%$	$n_{2DD}\%$

Figure 5: Inventive-step analysis of IMRaD over prior literature.

Table 1. Where and how to cite prior literature in new paper.

If	Then	Note
M of your **IMRaD** overlaps with prior literature's **M**	Your **M** may cite the prior literature	The prior literature and your paper adopt a common research method.
I of your **IMRaD** overlaps with prior literature's **I**	Your **I** may cite the prior literature	The prior literature and your paper have a common research background.
I of your **IMRaD** overlaps with prior literature's **D, R**	Your **I** should cite the prior literature	The prior literature is a direct prior literature of your paper.
D of your **IMRaD** overlaps with prior literature's **I, D, R**	Your **D** should cite the prior literature	You need to explain what makes your **D** different from **I, D, R** in the prior literature.
R of your **IMRaD** overlaps with prior literature's **I, D, R**	Your **D** should cite the prior literature	You need to explain what makes your **R** different from **I, D, R** in the prior literature.

In summary, academic paper writing can be assisted through AI, and the proposed 6-step model of academic paper writing (Figure 4) is the answer to RQ1: an AI specialized in natural language processing can streamline the process of writing academic papers, especially the process of prior literature review.

DISCUSSION: FUTURE OF RESEARCH PROFESSION IN AI ERA

In answering RQ1, we have shown that prior literature review can be streamlined through AI. In particular, the literature-analysis portion can be streamlined on a large scale through AI. In manual prior literature search, it takes a long time to find hit papers. There is also the risk of excluding important papers due to how much a person can read. However, with AI, the search can be done efficiently and comprehensively. Moreover, papers found with the help of AI in Step 2 are those in line with the researcher's interests, i.e., hit papers. Hence, researchers expect to be able to learn more efficiently about the RQs they have; they no longer have to search for one or two papers from hundreds or thousands. Therefore, we answer RQ2 in the affirmative: the researcher's ability to write academic papers will be enhanced with the assistance of AI (Kohda (2020, 2022)).

With some degree of precision, we discuss what will happen to the research profession in the AI era on the basis of the results of this study. If we can leverage Step 6 of the 6-step model of academic paper writing (Figure 4), the time that researchers pay for writing academic papers would be reduced by far. If one can come up with the rough outline of the IMRaD of a new paper, the AI will determine the originality and academic significance and even find prior literature that should be cited in the new paper. If so, is there no longer a need to do manual prior literature review? However, this discussion includes one large assumption: if one wants to write a paper that is original and academically significant, they must come up with a new original and academically significant IMRaD. Identifying such a unique IMRaD will require human curiosity and intuition. Until AI develops the equivalent of human curiosity and intuition, the task of looking for such IMRaDs, i.e., Steps 2 to 5 of the proposed 6-step model (Figure 4), must be human-driven. Hence, the research profession will not disappear while AI remains as it is.

CONCLUSION

We discussed academic paper writing in the AI era, arguing that AI-assisted academic paper writing is feasible since a service for AI-assisted patent prior art search is currently available (AI Samurai, 2019). We also argued that the use of AI would facilitate the development of researchers' abilities by letting them read hit papers that match their interests.

Shon (2015) made an interesting claim: students cannot write academic papers because they have never been taught how to read academic papers. He then proposed a set of analytical codes that helps IMRaD reading of social science papers. We hope that AI will be developed to assist in reading academic papers, for example, by using Shon's analytical codes. If so, the development of researchers' abilities will be further enhanced through AI.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP21K02829.

REFERENCES

AI Samurai. (2019). AI Samurai® [AI Patent Literature Evaluation System]. https://aisamurai.co.jp/

Frey, C. B., Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. Technological forecasting and social change, 114, 254–280.

Kohda, Y. (2020). Can humans learn from AI? a fundamental question in knowledge science in the AI era. AHFE 2020. Advances in intelligent systems and computing, vol. 1208. Springer, Cham. https://doi.org/10.1007/978-3-030-51057-2_34

Kohda, Y. (2022). A critical knowledge management question in the artificial intelligence era: "can humans learn from artificial intelligence or not?". Chandos Publishing, https://www.amazon.co.jp/Boosting-Knowledge-Economy-Environments-Professional/dp/1843347725.

Kohda, Y., Endo, S. (2006). Information retrieving apparatus and system for displaying information with incorporated advertising information. U. S. Patent No. 7, 136, 853. Washington, DC: U. S. Patent and Trademark Office.

Kohda, Y., Shirasaka, H. (2021). Patent evaluation and challenges with artificial Intelligence. Technical Information Institute Co. Ltd. ISBN: 978-4-86104-979-8, 44–56 (in Japanese).

Nair, P. R., Nair, V. D. (2014). Organization of a research paper: The IMRAD format. Scientific writing and communication in agriculture and natural resources, 13–25.

OpenAI. (2022). ChatGPT [Large language model]. https://chat.openai.com

Shon, P. C. (2015). How to read journal articles in the social sciences: A very practical guide for students. SAGE Publications Ltd; Second edition, 144 pages.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of business research, 104, 333–339.