Use of AI for Manuscript Writing – A Study Based on Patent Literature

Dr Shabib Ahmed Shaikh; Dr Nishad A Deshpande; Alok Khode
CSIR Unit for Research and Development of Information Products (URDIP), Pune

***Abstract* – Manuscript helps disseminate research output via conferences and publications. Writing research articles involves lots of time along with sound knowledge of the topic in question. However, in recent time, the shorter notice periods of manuscript submissions in conferences as well as in special issues from publication houses, has led to researchers looking for easier, simpler and faster techniques for writing research articles. As with all domains, technology is similarly helping researchers to make use of artificial intelligence for faster and plagiarism free writing of manuscripts with minimum human intervention.**

Even though, there is a debate on ethics of the usage of such technologies towards auto writing of manuscripts, the literature present in this work highlight that there is wide spread use of such tools and applications in manuscript writing. Without discussion on the advantages and disadvantages of artificial intelligence for manuscript writing, this paper makes use of patent literature available in public domain in the field of manuscript writing using artificial intelligence to understand the trends, techniques and technologies being used in this domain.

***Index Terms* – Artificial Intelligence, Manuscript Writing, Patent Analytics, Manuscript Generation**

1. INTRODUCTION

Manuscript writing is an important aspect of research. It represents a researcher's work in a standard, relevant, concise, fast, effective and scientific manner. A manuscript encourages exchange of ideas globally, receives recognition for the research as well as gets a work to be seen by the research community (Springer 2020). Writing a manuscript of a research work is one of the most time-consuming process requiring rigorous and repetitive checks. Though manuscript writing has different styles and reporting format based on predefined requirements as well as the content in question, the basic underlying objective of reporting a research outcome remain the same. Manuscript writing has many facets and require a deep understanding of the subject in question. Avoidance of repetitive ideas / theories as well as simple and easy explanation of a work for understanding of all is the key to any successful manuscript writing. Researchers and writers have always yearned for easier and simpler ways to write articles and looked forward for techniques and tools that make their output dissemination faster as well as striking compared to the conventional techniques. To understand availability of such techniques, this paper looks at the availability of artificial intelligence technology used for automatically writing of full or partial manuscripts.

It is important to note that this paper does not discuss the pros and cons of making use of AI in auto content generation nor do the authors encourage or discourage the use of such automatic manuscript writing applications and also do not endorse any of the applications mentioned in this work.

Artificial Intelligence (AI) enables a computer to think as human without any human intervention. Artificial Intelligence's disruptive technology has transformed several industries by automating processes in effective and efficient ways. AI has made its mark on all technological domains and manuscript writing is not optional. AI-based applications are being developed and implemented not only to generate scientific content but also to expedite the scientific communication as well as reduce human intervention (Enago Academy).

Manuscripts usually follow a rigidly defined structure and focus on a singular key message that logically and structurally support an idea in question that is communicated in the title (Gewin, 2019). Artificial Intelligence makes use of this basic rule to form the basis of automatically writing of manuscripts either fully or partially.

To understand the techniques and technologies used by artificial intelligence for writing of manuscripts, this paper analyses the patent data available in the said domain. Patents are techno-legal documents highlighting the legal as well as technological aspect of any scientific work (Deshpande et al., 2016; Shaikh & Londhe, 2016). Proper understanding and analysis of the technological aspect of the patent literature can highlight various aspects of a technology including its use and implementation (Shaikh & Singhal, 2019).

2. LITERATURE AND SOFTWARE APPLICATIONS RELATED TO USE OF AI IN PUBLICATION

The first modern scientific research paper was published in 1665 (Liumbruno et. al, 2013). Much has changed since then in the way articles are written, submitted, peer-reviewed, approved and published. Though the basic reporting matter remains the same, new standards and rules are in place for better ease towards submission of manuscripts in targeted journals of specific publishing houses.

Literature suggests availability of many online applications to help researchers write manuscripts that make use of Artificial Intelligence. "sciNote Manuscript Writer" is one example of such application that makes use of AI to prepare a draft of a scientific manuscript based on the data provided (Pavlek, 2017). It structures data into a series of tasks that contain the required data for manuscript preparation. An end user can write notes, add text, checklists, pictures, tables and upload different types of files within each task. Tasks can be used to define each step of a protocol in great detail, what needs to be done and how. Based on the data supplied in SciNote, it generates a draft manuscript containing several versions of the Introduction, Materials and Methods, Results and References.

Another example is the program named “SCIgen” that was created by scientists at the Massachusetts Institute of Technology “that generates random Computer Science research papers, including graphs, figures, and citations. It uses a hand-written context-free grammar to form all elements of the papers” (SCIgenWebsite).

EssayBot is another writing assistant powered by Artificial Intelligence (AI) that generates parapharase sentences, searches for data source as well as auto completes sentences using AI (EssayBot Website). The best part of EssayBot is that it even paraphrases the document concerns related to plagiarism checks and also generates citations matching the document.

EssayTyper is another example that makes use of some patented combination of algorithms and Wikipedia to develop content. The website tells that it helps one write manuscripts faster but also warns users that this should not be used legitimately as it can lead to plagiarism (EssayTyper Website).

AI Writer is another application that claims to help authors save 33% of their time. It accepts a headline or few keywords to generate unique and information-dense article written for the end user. Authors can re-use an already written articles by using a text spinner that rewords the contents (AI Writer Website). Same is the case with the Zyro AI content generator that creates technical texts which is unique by analysing hand-picked copy from each industry (Zyro Website). InferKit claims to be a State-of-the-art text generation that offers an interface and API for custom AI–based text generators for writers as well as developers (InferKit Website). Another example is textengine.io which independently generate individual texts using AI’s Natural Language Generation (NLG) to transform structured data into meaningful text (TextEngine Website). AutoSummarizer.com is another online AI based tool which can be used for Multi-Language Summarizing as well as paraphrasing (AutoSummarizer website). The online TEXT GENERATOR available at https://www.artikelschreiber.com/en/ is a free SaaS based text generator that makes use of English AI based Article Writer in three simple steps (ArtikelSchreiber website). Many text generation APIs are also available as opensource backed by a large-scale unsupervised language model that can generate paragraphs of text. Such transformer-based language model are based on the GPT-2 model by OpenAI which accepts a full or partial sentence to predict subsequent text from that input (DeepAI website). Articoolo is another application that helps writers with creating unique textual content from scratch, simulating a human writer by choosing the topic and length. It also helps paraphrase the already written article (Articoolo website).

Writing articles using Artificial intelligence is also playing an important role in journalism which is also referred to as automated or algorithmic or robotic journalism, in which AI enabled software are producing news articles (Carlson, 2015 ; Graefe, 2016). Software apps and algorithms are being provided by data science and AI companies to news outlets and media organizations have already started making use of the robotic journalism (Wölker & Powell, 2018 ; Fanta, 2017). Such automated articles are being widely used in the field of finance and sports (Fanta, 2017).

Many researchers have also discussed the pros and cons of using automated AI based tools for manuscript preparation. While many feel that such tools and applications help faster dissemination of knowledge, researchers have also raised alarm on junk content creations and their communications in scientific and research conferences by making use of algorithms which do not make any sense for readers (Bohannon, 2013; Labbé, 2016; Graefe, 2016; Fanta, 2017). Such AI based tools and programs have found users across the globe and the manuscripts generated are being accepted by scientific conferences and published in purportedly peer-reviewed journals leading to retraction of 122 such papers by Springer and IEEE. The publication

house Springer had announced a software "SciDetect", a program that automatically detect papers created with automated tools to "ensure that unfair methods and quick cheats do not go unnoticed" (Bohannon, 2015). This has led to call for higher standards of peer review of contents submitted in conferences and publications.

There have been reported instances where programs were used to produce such automated manuscripts and were accepted and published in predatory journals, however it was found out that such manuscripts were not meaningful and were just concoction of some technological words that did not convey any meaning. Wiley in-fact has gone ahead and has been using a software program to detect manuscripts prepared by computers. Also recently, The Guardian published an article that was generated by AI based program (The Guardian, 2020). However, to make it readable it had to be edited by humans. Therefore, it can be safely assumed that the present AI lead manuscript writing programs cannot replace the human intellect, although humans cannot rule out possibility of becoming indispensable.

3. RESEARCH METHODOLOGY

For carrying out the searching and extracting patent data related to the use of AI for article writing, the Derwent Innovation Database (www.derveninnovation.com) was used. The retrieved patent data was accordingly cleaned, normalised and then analysed using excel spread sheet. The methodology involved the following two major steps as outlined below.

Information gathering:
A search strategy was prepared and then executed on the Derwent Innovation commercially available database, in order to find applied and granted patents in the field of AI for article writing. Patents applied or published till October 31, 2020 were considered in the study. The search strategy consisted of international patent classification (IPC) codes (EPO, 2016) relevant to the topic as well as keywords.

The search strategy yielded ~900 patent records from the Derwent Innovation Database. This included all patent application published as of October 31, 2020. As patents are territorial in nature, the same invention may be duplicated by way of multiple filings in different countries, known as patent families (EPO, 2016). To reduce this form of duplication, one representative for each patent family was retained to obtain a dataset of ~290 patent records.

The bibliographic details of patents such as the title, abstract, claims, priority date, assignee name, inventor names, INPADOC family members, citations, etc. were collected and stored in a spreadsheet for further processing and analysis.

Data cleaning:
Before analysing the data, it was necessary to normalize it. There was also a need to take into consideration the spellings and abbreviations of names of assignees apart from spellings & typographical errors, mergers, acquisitions and corporate hierarchies of various assignee organizations. Hence the patent application data was normalized for assignee names by replacing name variants with a standard name. The inventor list was also cleaned and Assignee normalization was also carried out.

4. DATA ANALYSIS AND VISUALIZATION

After cleaning and normalization of data, the tittle, abstract and claims were analyzed to understand the technologies highlighted in each of the patents which directly represented the use of AI for article writing. After reading, it was found that 48 patent applications were relevant to the study. Hence for further analyzing and visualization, these 48 relevant patent records were considered. The analysis was carried out using spread sheet and by reading and understanding individual patents based on technology.

4.1. Patenting activity

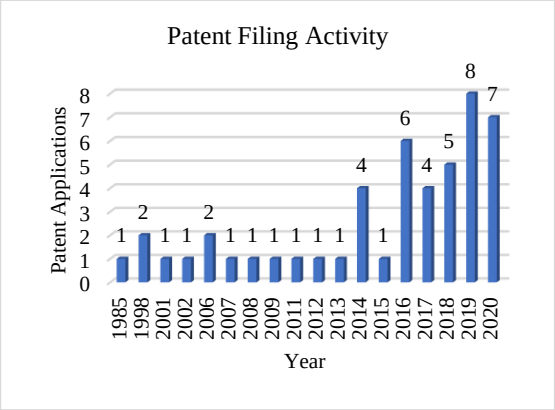

Figure 1. Over all patent filing activity.

Figure 1 above depicts the overall patent filing activity. It is observed that the earliest patent relating to use of AI powered manuscript writing was filed in the year 1985. The patent filing activity has picked up from year 2016 and is on rise ever-since. A patent gets published in public domain only after a period of 18 months from its application, hence the patenting activity reported above for the resent years might not be the actual case as many patent applications might not have been published in the last two years (Shaikh et al., 2018). The first patent published in 1985 belonged to IBM and was used for "compaction and replacement of phrases" and it was protected in US, Canada, Europe and Japan. An

important point that came out from the patent literature was that initially all the patents using this technology were for paraphrasing or small sentence formulation. The first patent filing from China was in 2012 for checking of plagiarism.

4.2. Application Filing Jurisdiction:

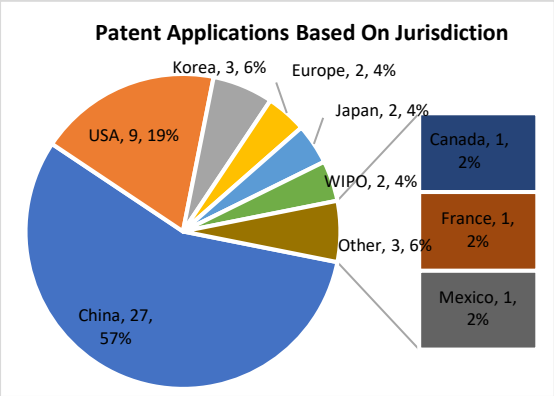

Figure 2. Jurisdiction based patent filing activity.

The figure 2 above represents the jurisdictions in which patents for AI based manuscripts were filed. It can be observed that 67% of the filings on this technology is from Asian subcontinent. China leads with 57% of the patents in this field wile USA has a filings of 16%.

Due to territorial nature of patent protection, separate applications for same inventions are filed in different countries and such sets of related applications are called patent families (Patil et al., 2019). To reduce the duplicity brought in by multiple filings, in the present analysis, one representative member per family is considered. European Patent Organization provides grouping of family members called as INPADOC patent family (EPO) and in the present study, we have used the same while analyzing all the jurisdictions in which the patents for technologies were filed. It is highlighted in the table 1 below.

Table 1: INPIDOC Family members of patents being filed

Countries	INPADOC Count	Patent Applications	Average Filings
Canada	1	1	1.00
China	51	27	1.89
Europe	15	2	7.50
France	1	1	1.00
Japan	3	2	1.50
Korea	5	3	1.67
Mexico	22	1	22.00
USA	71	9	7.89
WIPO	16	2	8.00
Total	**185**	**48**	3.85

It was observed that patents originating from USA are filed on an average in about 7.9 jurisdictions while china was having an average of 1.89 filings. One patent from China had 19 INPADOC family members while there were 20 patents which had only one INPADOC family member for patent application from China. One patent filed from Mexico was having 22 family members. Two patents filed from Europe had an average filing in 7.5 jurisdictions. 2 patents filed from WIPO had an average of 8 family patents

Table 2 : Year wise filings by countries

Year	Countries									Total
	CA	CN	EP	FR	JP	KR	MX	US	WO	
1985								1		1
1998			1					1		2
2001									1	1
2002							1			1
2006						1		1		2
2007				1						1
2008						1				1
2009								1		1
2011								1		1
2012		1								1
2013								1		1
2014	1	1						2		4
2015		1								1
2016		5			1					6
2017		2	1						1	4
2018		4			1					5
2019		6				1		1		8
2020		7								7
Total	**1**	**27**	**2**	**1**	**2**	**3**	**1**	**9**	**2**	**48**

It can be observed from the table 2 above that initially the patent filings in the domain of AI for manuscript writing was from USA and Korea. China entered into the foray in 2012 and its 24 patents out of 27 are filed in the last five years.

Even though patents were filed in the countries mentioned above, the assignee countries yielded a different picture. The details are highlighted in the table 3 below.

Table 3 : Assignee Countries from which the patents were filed

Year	Countries						Total
	AU	CA	CN	JP	KR	US	
1985						1	1
1998				2			2
2001				1			1
2002						1	1
2006					1	1	2
2007					1		1
2008					1		1
2009				1			1
2011						1	1
2012			1				1
2013				1			1
2014	1	1		1		1	4
2015			1				1
2016			5	1			6
2017			3			1	4
2018			4	1			5
2019			6	1	1		8
2020			6	1			7
Total	**1**	**1**	**26**	**10**	**4**	**6**	**48**

It can be observed that the initial players were from US, Japan and Korea only, even though China is leading the pack with 26 patents assigned. What is important to know is Japan has 10 patents and it has filed most of its patents in US and other territories.

4.3. Analysis based on technology and application

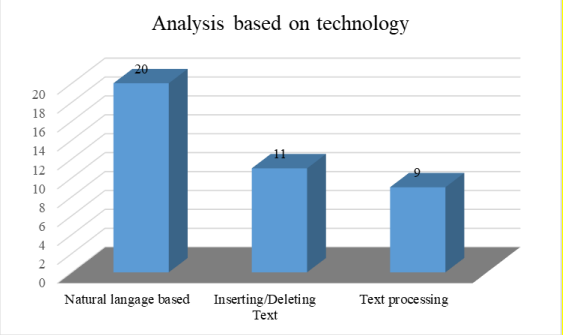

Figure 3 analysis based on top technology areas

Chart 3 above reveals major technologies involved. These major technologies were created by understanding the classification codes of the top 10 international classification codes (IPC) and then assigning the technological area for the patent. As can be observed, major technology involves use of Natural Language analysis and processing for Orthographic correction for text that includes spell checking or vowellisation (These were identified from the IPC codes G06F17/27and G06F17/30). These were followed by patents that were assigned IPC code of G06F17/24 and these are related to editing of the manuscript that involved insertion or deletion of the text which could mostly involve rephrasing. Next major group of technology involved patents related to IPC code G06F 17/21 that represented Text processing with a focus on the formatting aspects such as font handling, mathematical formulae representation etc.

Each of the patents were manually read to understand their claims and understand whether the patent represented composition processing or full generation of the manuscript. The details of which are highlighted below in figure 4.

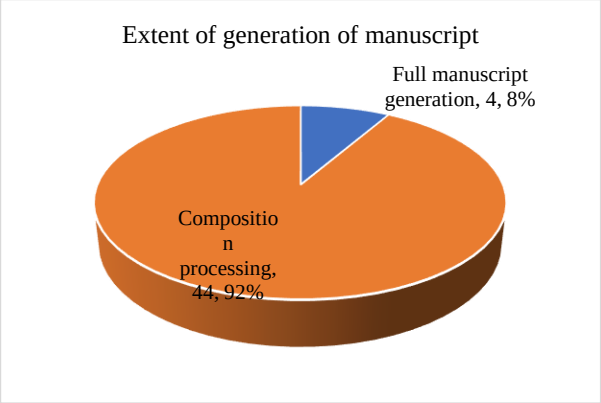

Figure 4 the extent of generation of manuscript

Figure 4 above reveals the extent to which the patents relating to AI assisted manuscript generation. It is observed that major patents assist in construction of the sentences that go into composition of the manuscript. Hence it can be said the major patents were related to paraphrasing of manuscripts. However, there are also few patents which claim automatic generation of the full manuscript.

4.4. Analysis based on the contents of the manuscript.

The patent data was further analyzed to understand what type of data was being paraphrased or getting modified. The details are given in figure 5 below.

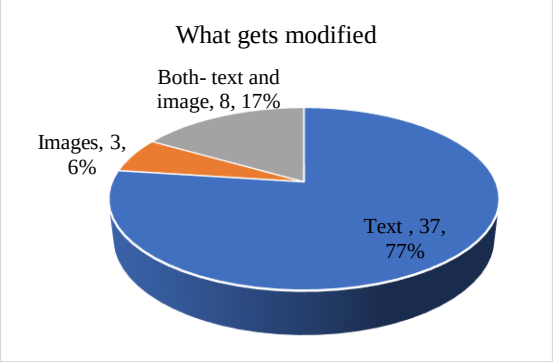

Figure 5 use of AI in manuscript preparation

The chart 5 above represents use of AI in preparation of specific portions of the manuscript such as text construction and labeling or positioning of images. It is observed that majority of the patents disclose use of AI for construction or modification of the text. There are few patents that claim use of AI for modification of both images as well as images.

5. DISCUSSION AND CONCLUSION:

Manuscript writing is a part of research that not only disseminates research outputs but also validates its results amongst the general public. It is one of the most complex and time-consuming tasks of a research and researchers have always yearned for faster and simpler ways of writing. Machine learning in general and Artificial intelligence in specific has come to the rescue of authors to automatically write manuscripts with minimal human intervention. Even though the use of such technologies goes beyond the basic ethics of paper writing, applications based on such technologies have found wide spread popularity amongst authors who look for quick and fast research writings. To understand the growth of these technologies, this paper looks at patent literature and highlights the trends of its technology.

It was observed from patent data that AI was not only used in auto manuscript writing but also used for paraphrasing as well as formatting the manuscript. The patent literature also highlights the techniques deployed and the major nations in the field of AI for manuscript writing. From the analysis it was observed that patents in the field of manuscript writing with the help of AI existed from the mid 1980's when the first patent was filed by IBM for parapharases. However 63% of the patents in this filed are filed in the past 5 years only. It can be observed from the analysis that the initial players in the field of manuscript writing using AI were from US, Japan and Korea only. However, such technologies are protected in all major nations and China has dominated this domain in the last 5 years. What is important to know is Japan has 10 patents and it has filed most of its patents in US and other territories. The first patent filed from China in this domain was in 2012 and it was for checking of plagiarism.

Patents from China are aggressively on a rise accounting for a total of 27 patents out of the total 48 patents in this area. 24 patents from China are filed in the last 5 years accounting for 50% of the total patents filed for manuscript writing using AI. Hence it can be concluded that 80% of the patents filed in the technology of manuscript writing using AI originated from China since 2016.

Most patents filed from China have only one INPADOC family member, showing that it is protected in only one country. Around 74% of the Chinese patents were having only one INPADOC family member and there also existed one Chinese patent with 19 INPADOC family members. On the other hand, a sole patent filed from Mexico had 22 family members. It was also observed that major technology involved the use of Natural Language analysis and processing for Orthographic correction for text that includes spell checking or vowellisation which pointed towards the use of technologies related to paraphrasing of manuscripts. Majority of the patents disclose the use of AI for construction or modification of the text. However, there were few patents which claim automatic generation of the full manuscript. Important point that came out from the patent literature was that initially all the patents using this technology were for paraphrasing or small sentence formulation.

The advances in the technology have resulted into use of AI for preparation of manuscript. Initially the use was sentence framing and for grammatical construction and correctness. As the technology refined, the patents that rephrase/modify the sentences based on the similarity were found that performed rephrasing to overcome similarity after plagiarism check. Recent filings also claim automatic generation of manuscripts. However, this is possible based on the training sets based on inputs models (existing manuscripts for the papers). An important point to note is that there is always something new in any research and a research manuscript always points to that newness contributed to the body of existing knowledge. On a positive note, AI's won't be able to replace the human factor involved in writing manuscripts as AI based technologies are dependent on existing historic data where as a research paper highlights new evidence. Even then, the authors neither encourage or discourage the use of such automatic article writing applications and do not in any way endorse any of the applications, technologies, algorithms etc. mentioned in this work.

6. REFERENCES

AI Writer Website. accessed online fromhttp://ai-writer.com/

Articoolo website. accessed online from http://articoolo.com/

ArtikelSchreiber.com website. accessed online from https://www.artikelschreiber.com/en/

AutoSummarizer website. accessed online from https://autosummarizer.com/

Bohannon, J. (2013). Who's afraid of peer review?.Science 342 (6154), 60-65

Bohannon, J. (2015). Hoax-detecting software spots fake papers. Scientific Community, Scientific Publishing. accessed online from https://www.sciencemag.org/news/2015/03/hoax-detecting-software-spots-fake-papers

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. Digital journalism, 3(3), 416-431.

DeepAI website. accessed online from https://deepai.org/machine-learning-model/text-generator

Deshpande, N., Ahmed, S. and Khode, A. (2016) Business intelligence through patinformatics: A study of energy efficient data centres using patent data. Journal of Intelligence Studies in Business. 6 (3) 13-26

EPO website. Accessed online from https://ie.espacenet.com/help?locale=en_IE&method=handleHelpTopic&topic=legalstatusqh

Enago Academy, NA, Should AI Write Research Papers on Behalf of Authors ?; accessed online from https://app.enago.com/homepage/i/30630/should-ai-write-research-papers-behalf-authors

Essaybot Website. accessed online from https://www.essaybot.com/

EssayTyper Website. accessed online from http://www.essaytyper.com/about

Fanta, A. (2017). Putting Europe's robots on the map: automated journalism in news agencies. Reuters Institute Fellowship Paper, 2017-09.

Gewin, V. (2018). How to write a first-class paper. Nature, 555(7694). Accessed online from https://www.nature.com/articles/d41586-018-02404-4.

Graefe, A. (2016). Guide to automated journalism. https://academiccommons.columbia.edu/doi/10.7916/D80G3XDJ

Graefe, A. (2016). Guide to automated journalism.Columbia Journalism School. Tow Center for Digital Journalism

InferKit website. accessed online from https://inferkit.com/

Labbé, C., Labbé, D., & Portet, F. (2016). Detection of Computer-Generated Papers in Scientific Literature. In Creativity and Universality in Language (pp. 123-141). Springer, Cham.

Patil, S. J., Bhalke, S. L., & Deshpande, N. (2019). A state of the art in genetic improvement of Asparagus plants: Patent based perspective. Journal of Commercial Biotechnology, 25(1), 8-23

Pavlek. T. (2017). SciNote Can Write a Draft of Your Scientific Manuscript Using Artificial Intelligence. accessed online from https://www.scinote.net/scinote-can-write-draft-scientific-manuscript-using-artificial-intelligence/

SCIgen Website. accessed online from https://pdos.csail.mit.edu/archive/scigen/

Shabib-Ahmed, Londhe, B. R. (2016). Intricacies of Software Protection: A Techno - Legal Review, Journal of Intellectual Property Rights, Vol 21, 157 - 165

Shaikh, S. A., & Singhal, T. K. (2018). An analysis of IP management strategies of ICT companies based on patent filings. Journal of Intelligence Studies in Business, 8(2).

Shaikh, S.A**.** and Singhal, T.K. (2019) Study on the various Intellectual Property Management Strategies used and implemented by ICT firms for Business Intelligence. Journal of Intelligence Studies in Business. 9 (2) 30-42.

Springer (2020), Writing a journal manuscript; accessed online from https://www.springer.com/gp/authors-editors/authorandreviewertutorials/writing-a-journal-manuscript/author-academy/10285484

TextEngine website. accessed online from https://www.textengine.io/en/

The Guardian (2020), A robot wrote this entire article. Are you scared yet, human? accessed online from https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

Wölker, A., & Powell, T. E. (2018). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. Journalism, 1464884918757072.

Zyro Website. AI Content Generator from zyro website accesses online from https://zyro.com/ai/content-generator

Author 1 : Shabib Ahmed Shaikh is working as a Senior Technical Officer at the CSIR-Unit for Research and Development of Information Products. He has more than 20 years of work experience in the field of Information Technology. He has a PhD in Intellectual Property Management Strategies. He is a graduate in Electronic Science and a Post Graduate in Computer Management. He was awarded a merit for M.Phil. He is actively working in the domain of Patent Analytics, Data Science, Open Access, Institutional Repositories' and their Harvesting

Author 2 : Nishad Deshpande is a Senior Scientist at CSIR Unit for Research and Development of Information, Pune, India. He is having more than 15 years of experience in Life sciences and as a patent analysist. He holds a PhD from Symbiosis International University and is an expert in the field of patinformatics. He has been awarded SIS young information Scientist by Society on Information sciences for his contribution towards data analytics and patinformatics.

Author 3 : Alok Khode works as Principal Scientist at CSIR - Unit for Research and Development of Information, Pune, India. He is having more than 20 year of experience. His research interests are mainly but not limited to information retrieval, knowledge management and open access. He has master degree in computer application from DAVV Indore and pgCert from Cranfield University, UK. He has been awarded Chevening Cyber Security Fellowship by UK goverment as well as SIS Fellowship by Society on Information sciences for his contribution towards open access movement in India.