Artificial Intelligence Is a Tool for Cheating Academic Integrity

El-Sayed Abd-Elaal, Sithara H.P.W. Gamage, Julie E. Mills
School of Natural and Built Environments, University of South Australia, Australia
Corresponding Author Email: abdee001@mymail.unisa.edu.au

Introduction

Jeremy Stribling and his colleagues established an automatic paper generator which cobbles parts of articles together with randomly generated graphs. They aimed to test whether such nonsense articles could pass the screening procedure of the 9th World Multi-Conference on Systemics. Unfortunately, the Cybernetics and Informatics (WMSCI) conference accepted those papers (Ball, 2005; Tien & Labbé, 2018). Van Noorden (2014) reported that, the publishers Springer and IEEE removed more than 120 papers from their subscription services after discovering that those papers were computer-generated nonsense. These kinds of papers can easily deceive inexperienced readers because they have proper sentences and structure as in a normal scientific paper.

Since the development of the first online article generator in 2005 by Stribling, Krohn, and Aguayo (2005), significant steps have been taken to further develop those article generators for useful purposes. As an example, Malviya and Tiwary (2016) developed a system using AI to generate a research article that summarised previous studies. Also, others used AI for developing tools to mark students' essays automatically and these tools are called "robo-graders" (Smith, 2018).

On the other hand, some people have misused these AI tools. For example, Ike Antkare who presented more than 100 publications in a year, which brought his name to the 21st position of the most cited scholars in the world, which is higher than Einstein (Labbé, 2010). Philip Parker wrote 200,000 books with the assistance of several computers and programmers, and he is selling these books on Amazon. He produces a book every twenty minutes using a patented process which can generate books automatically by using internet and database searches (Cohen, 2008; Spinak, 2014). Interestingly, most of these books do not have a cover page as AI is still a little bit behind achieving such a task.

Currently, there are a significant number of advanced Automatic Article Generators (AAGs) that can generate essays, articles, research papers, books and theses. Detecting a work generated by AAG creates extra burden on teachers, researchers and anti-plagiarism tools. The OpenAI (an AI research company) recently announced that they have built an advanced text generator which is considered too dangerous to release (Whittaker, 2019). According to the OpenAI, their new natural language model, GPT-2, was trained to predict the next word from a sample of 40 gigabytes of internet text, allowing users to generate realistic and coherent text for a chosen topic (Whittaker, 2019).

Hiney (2015) stated that fabrication, falsification and plagiarism are the core of research misconduct. Peterson (2007) stated that misconduct in research caused by fabrication and falsification destroys the scientific enterprise, undermines trust and is a misuse of public funds. Using AI tools such as text generators, paper generators or thesis generators is serious academic and research misconduct as it does not reflect the author's genuine work. As these AI tools beat the systems that detect academic and research misconduct such as plagiarism, detecting papers generated by AI needs a vigorous teacher/ reviewer.

In the past decade, ghost writing (getting someone else to do the writing), also known as the 'paper mill' approach was also an issue of academic integrity (Singh & Remenyi, 2016). However, this issue is now diminishing while article generation is vastly increasing. Google

trend analysis from 2004 to 2019 indicated that Australia received a score of 100 for the searches of both phrases “Ghost writing” and “Article generation”. This means the students in Australia have searched these two phrases more than in any other English speaking countries. Further investigation of Google trends also showed that the searched term “Ghost writing” was decreasing over the years of 2004-2019 by 75% while the searched term “Article generation” was increasing from 2004-2014 and then it started to decrease after 2014. Such a decline after 2014 could be due to the development of the article generator detection tools such as SciDetectTM. According to Tien and Labbé (2015) SciDetectTM detected more than 120 phony articles published through Springer and IEEE.

In this paper, the authors explain the concepts and techniques behind the AAGs, then the possible approaches for identifying the generated phony texts and their characteristics. Later, the related fabrication and plagiarism issues associated with AAGs are discussed. Finally, the teachers’, supervisors’ and institutions’ roles to mitigate this potential academic and research misconduct are outlined.

Automatic Article Generator (AAG)

AAG is an online computer software that has access to a large database including old and new academic articles, books and reports. These computer programs search the databases according to user given keywords, or the suggested research topic and then they compose related articles in a requested style. These generators utilise jargon (jargon consists of particular terminologies which are related to a specific research area and may not be well understood outside of that context) from the related field to compose sentences. Kelly-Bootle (2005) indicated that the non-sophisticated reader could be deceived by the effective use of jargon. These programs use context free grammar tools (a set of grammar rules used to produce patterns of strings in a formal language). By using Context-Free Grammar tools with features to replace some of the articles’ contents with random synonyms, the resultant writing will not be a direct copy from published work and therefore plagiarism detection tools cannot detect this fabrication.

Malviya and Tiwary (2016) described an advanced approach for ‘Summarised Research Article Generator’ which can be used to generate summarised research papers. The approach searches the related articles first and then extracts and stores the semantic information from those articles in a knowledge tree. Each extracted textual unit will be stored with a score, so textual units will be arranged at different levels. For example, the words used, along with the probability, expressions and phrases beside their paragraphs and sections will be stored with their scores. Secondly, the scores of all the units will be calculated. Finally, Bayesian network theory (Bayesian Network refers to random variables and conditional dependencies) will be utilised to develop a probabilistic model to extract the relevant information from the tree to generate the article. They also suggested that articles should be developed in segments by combining the most relevant paragraphs for each section separately such as abstract, introductions, methodology, and so on as shown in Figure 1.

Another big development is ‘generating natural language techniques’ which develops wording very close to human-crafted phrases (Nguyen-Son, Tieu, Nguyen, Yamagishi, & Zen, 2017). However, according to the description given by Mugan (2018) the developed computers conceptual systems are not rich enough to represent what a little child is aware of. Similarly, AAG’s tools could be good for collecting similar phrases from similar paragraphs and related publications but these tools are not aware of the meaning of each phrase they collect. Therefore, AAG’s product is not coherent on a large scale and it often doesn’t make sense even though it might be rich with information. By contrast, recent researchers developed a method that encodes representations of our world to a space of meanings, then transforms points from that space into a language (Mugan, 2018). Potentially these researches may help AAGs to generate logical and much more meaningful content however, this kind of research is still at its early stage.

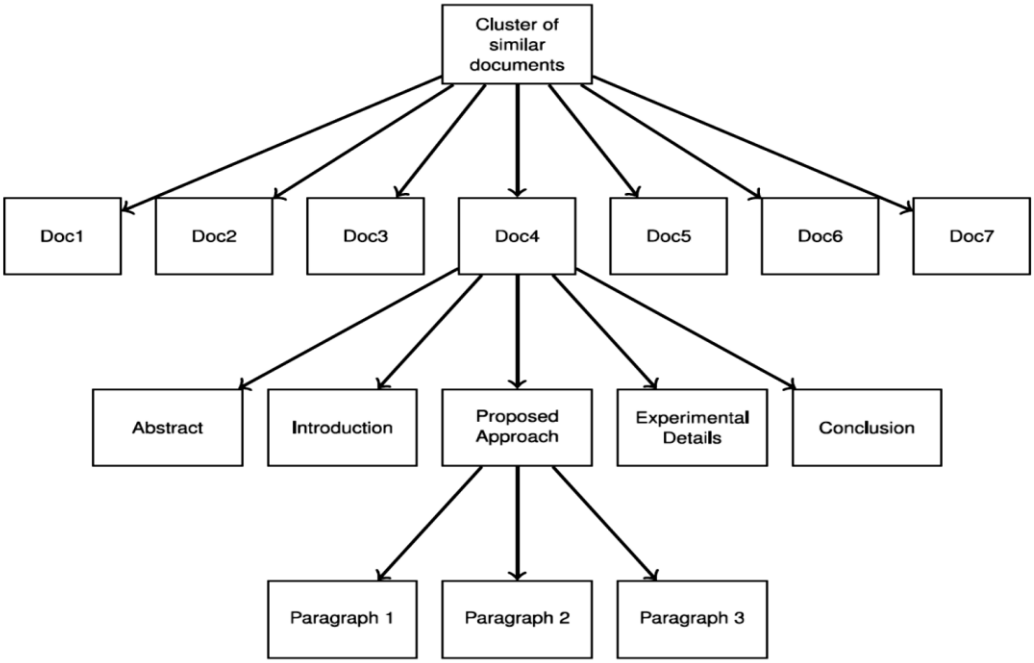

Figure 1: Structure of Knowledge Tree (Malviya & Tiwary, 2016).

In this study, the authors used one writing generator tool, the Article Generator Pro ("Article Generator Pro,"), which writes a part of the paragraph free of charge. It needs user inputs such as keywords, word limit and reference style, then it generates some writing. In this paper, the authors used the keywords “Artificial intelligence tool cheating academic integrity” to generate part of a paragraph and four outputs shown in Figure 2.

The full generated articles have a realistic structure, a set of references, reasonable figures and format including abstract, introduction, methodology, conclusion and reference list. Mathgen (Eldredge, 2012) and SCIgen (Stribling et al., 2005) can generate full mathematics research papers or computer science research papers, respectively. Both are free tools and can generate hundreds of papers in few hours, however both generators are not sophisticated enough to generate articles using specific keywords.

“Why did we go down this path? When we analyzed cheating behavior at a macro level, we found that nearly every behavior has a pattern. From the beginning, we had been training our human proctors to review those specific actions. Since we can train a human to recognize these patterns of manner, we knew we could train a machine to do the same. However, in our live proctoring breach, technology is interest to enhance earthli..”
“From a pure user experience standpoint, the problems are remarkably similar. With essay mills a student (or other wannabe author) can input a topic plus a few paramaters and get a fully-written work. The only difference is that, on the backend, it’s a human generating it not an AI. When we look back on writing papers for school, at some point, we have all experienced the feeling that the work is non-transfe...”
Though the idea of robots writing multitude papers might seem to be the field of science fiction, the truth is robots are already writing content. In September 2017, the Washington Post announced their AI reporter, dubbed Heliograf, had composed some 850 stories in the preceding year. This inclosed some 300 reports from the Rio Olympics.
Even top tertiary institutions like Harvard, Stanford University, and Duke University among others have not escaped cheating scandals. According to the International Center for Academic Integrity (ICAI), about 68 % of undergraduate students and about 43 % of graduate students admit to cheating on tests or in written assignments. Turnitin’s various products are in wide use: more than 30 million students from more than 15...

Figure 2: Robot writing examples

Identifying Automatically Generated Papers

Several approaches have been identified to detect computer generated papers. Xiong and Huang (2009) noted that most of the references included in the papers generated by SCIgen (Stribling et al., 2005) are fake and do not exist in the Internet. However, the subsequent versions of AAG's tools had handled this defect. Amancio (2015) suggested a method to compare the topological properties of complex networks (machine writing network) of artificially generated manuscripts by SCIGen against authentic papers. Also, Tien and Labbé (2018) investigated the similarity in grammatical structure of sentences to detect the sentences that are created using a Probabilistic Context-Free Grammar software. Although, this approach helped to catch many gibberish papers, in particular the second one caught more than 120 fake published articles, they are not effective against article generators that use other writing techniques such as Recurrent Neural Network or Markov chains.

Recently, Gehrmann, Strobelt, and Rush (2019) developed a tool called GLTR (A tool to detect automatically generated text) by conducting statistical analysis for each word in the generated text. The tool conducts three tests, the probability of the word, the rank of a word and the entropy of the predicted distribution. Adelani et al. (2019) investigated fake online reviews and found that GLTR can accurately detect fake reviews. However, in this current study, the authors tested the generated paragraphs (shown in Figure 2) using GLTR. Unfortunately, this tool indicated those generated paragraphs as human writing.

Lavoie and Krishnamoorthy (2010) concluded that the keywords that appear in the title and abstract should appear more frequently in the body of human written papers. In addition, they mentioned that, the human articles should have keywords from the title of the cited articles. Furthermore, certain keywords might be preferred by the author throughout a paper. That suggestion agrees well with what is known by author personal's tic (tic' means some specific words that author tends to use them more than usual, such words "also", "significant", "highly important", "however", "could", "agree", "Unfortunately" etc). However, these approaches may not be significant for a short piece of writing.

The authors believe that the Lavoie and Krishnamoorthy (2010) approach (investigating patterns of keywords and overused words), is meaningful because AAG cannot easily create those keyword patterns, despite developers being aware of it. Also, the investigation of keywords should be extended into the conclusion section. In addition, investigating carefully the cited reference list will help to identify whether it is a generated article or not. AAG may copy from articles with similar titles to the researched topic, but those cited articles may discuss subjects that are not close to the actual topic.

Investigating the statistics of the cited journals is another way of detecting AAG. In a genuine paper, authors usually focus on a limited number of journals while AAG tends to use high numbers of diverse resources (e.g. articles could be from all over the world for no reason, or from many internet resources). Added to this, AAG's writing could include words that give different meanings based on a context, but not a clear message with a firm focus.

Potential Academic Integrity issues and solutions

The AAGs are currently available to anyone at an affordable cost and Google trends point to increasing searches for article generator tools. Therefore, there is a high possibility that some students are using AAGs for their academic work. As it can fool markers, students may be given credit for writing which does not reflect their true ability. Therefore, it is important to identify and come up with possible solutions for this new form of academic misconduct.

To overcome the existing and potential academic integrity issues associated with the article generators a number of steps can be followed. Firstly, it is vital to raise awareness among educators about the issues associated with these tools, because many educators may not be aware of this issue. One study found that the markers can often detect contract cheating when asked to do so, and training may be helpful in improving their accuracy of detecting the

cheating (Williams & Nash, 2009). The same study suggested that markers' suspicions are a crucial factor in addressing contract cheating. Therefore, raising the awareness of AI, and training the markers to identify such articles are very important aspects.

Secondly, it is important to conduct workshops to provide clear understanding of how AAG works and explaining the characteristics of the generated writing. Therefore, if a marker notices any features of generated writing, such as a significant number of references that a student has included in his/her work but the marker suspects that there was not enough time or possibility for the student to actually gather and read these, then it should raise suspicions. If a student's work has flawless grammar and is formatted incredibly well while the fundamental ideas are not there or don't make sense, then it should raise alarm.

Thirdly, educators need to understand the underlying reasons for plagiarism and should address these by providing extra support. The primary approach to address this issue, in the authors' opinion, is a more humanitarian approach. This may require several personal and institutional changes. One study has highlighted five conditions that cause academic dishonesty. They are academic aptitude, perseverance, personal issues, competing objectives, and self-discipline (Amigud & Lancaster, 2019). Some scholars suggest that learners who come from certain countries have a more accepting and permissive attitude towards cheating at universities (Ehrich, Howard, Mu, & Bokosmaty, 2016). Another study found that the main reason why students cheated was their workload (Costley, 2019). Researchers have pointed out various other reasons for why students/researchers take the path of academic misconduct and some researchers have come up with many strategies to prevent it (Awasthi, 2019). If the underlying reasons for academic misconduct are not addressed, people will try to beat the anti-plagiarism or anti article generator tools by coming up with even further advanced software packages.

Fourthly, it is also important to understand whether students realise that they are committing an offence of academic misconduct. A recent study found that a large percentage of students who study face-to-face courses and online courses did not consider their behaviours to be a violation of academic misconduct (Burgason, Sefiha, & Briggs, 2019). Therefore, it is important to educate learners and provide the necessary support they need so that they do not feel like there is a reason to cheat. However, conducting a direct discussion with students regarding AAG tools is a sensitive issue, as many students do not know these tools exist so such a discussion could unintentionally advertise them. A study conducted to investigate perceptions and motivations for cheating suggested that the general apathy students may have towards study causes cheating, and online classes increased that apathy (Costley, 2019). Widespread cheating among peers and acceptance by instructors also encouraged students to cheat (Costley, 2019). This is an important fact to take into consideration. If markers cannot detect automatically generated articles, then they unintentionally encourage other people to use it, so it is very important to address these issues.

Fifthly, giving students less opportunity to cheat is also a possible solution. This can be achieved by structuring the assessment types that are less likely to result in cheating. A recent study found that the assessment types such as 'review articles', 'literature analysis and thinking', 'integrated knowledge/skills', 'real word tasks' and 'no right answer tasks' are more likely to be out sourced by students (Bretag et al., 2019). These assessment types are, in fact, easy to produce using article generators. In contrast, a 'series of small graded tasks', 'in-class activities', 'personalised assignment' and 'customised practicals', 'interviews/vivas' are less likely to result in cheating. Therefore, enhancing the curriculum with assessments which are less likely to result in cheating can be a solution to article generation as well as some other forms of academic misconduct.

Finally, instigating severe penalties for offenders can also reduce the number of academic misconduct cases. However, this should not be the primary focus of educators, as this will not address the reasons why people are using these unauthorised services and committing academic misconduct. More research to develop tools to identify the characteristics of

automatically generated papers and flag them for the attention of the human graders is also a worthwhile investment.

Conclusion

This paper highlights the features of AAGs driven by AI. It also discusses the possible academic integrity issues associated with AAGs and possible solutions to those issues. The paper contains a description of the historic development of AAGs and examples of how people have utilised these AAGs to produce a large number of fake publications, some of which were even published in ranked journals. Then, the paper explains the techniques behind the AAGs, and the specifications and features of automatically generated writings. Finally, the responsibilities and potential techniques that academics and researchers could use to mitigate the risk of possible academic misconduct were also discussed.

The study, concluded that:

- AAGs are available to anyone, growing in numbers and rapidly improving in quality of their outputs.
- Although AAGs are powerful enough to collect the related critical information better than good researchers, these tools are not aware of the meaning of the collected information. Therefore, such writings may not make sense, logic or show clear cohesion between paragraphs.
- There are some approaches for detecting such fake writing, however, they need vigorous academic evaluation.
- The educators should be aware of AAGs and training workshops should be conducted to improve the chances of detecting fake writing.
- The academics/reviewers should be qualified to investigate phony articles and institutions should decide the appropriate punishments for the offenders.
- The teaching and research institutions should improve their plagiarism and fabrication policies against AI tools.

References

Adelani, D. I., Mai, H., Fang, F., Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Generating Sentiment-Preserving Fake Online Reviews Using Neural Language Models and Their Human-and Machine-based Detection. *arXiv preprint arXiv:1907.09177*.

Amancio, D. R. (2015). Comparing the topological properties of real and artificially generated scientific manuscripts. *Scientometrics, 105*(3), 1763-1779.

Amigud, A., & Lancaster, T. (2019). 246 reasons to cheat: An analysis of students' reasons for seeking to outsource academic work. *Computers & Education, 134*, 98-107.

Article Generator Pro. Retrieved 09/08/2019, 2019, from https://www.articlegeneratorpro.com/

Awasthi, S. (2019). Plagiarism and Academic Misconduct A Systematic Review. *DESIDOC Journal of Library & Information Technology, 39*(2), 94-100.

Ball, P. (2005). Computer conference welcomes gobbledegook paper. *Nature, 434*(7036), 946-946.

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., van Haeringen, K., et al. (2019). Contract cheating and assessment design: exploring the relationship. *Assessment & Evaluation in Higher Education, 44*(5), 676-691.

Burgason, K. A., Sefiha, O., & Briggs, L. (2019). Cheating is in the Eye of the Beholder: an Evolving Understanding of Academic Misconduct. *Innovative Higher Education, 44*(3), 203-218.

Cohen, N. (2008). He wrote 200,000 books (but computers did some of the work). *New York Times, 10*.

Costley, J. (2019). Student Perceptions of Academic Dishonesty at a Cyber-University in South Korea. *Journal of Academic Ethics*, 1-13.

Ehrich, J., Howard, S. J., Mu, C., & Bokosmaty, S. (2016). A comparison of Chinese and Australian university students' attitudes towards plagiarism. *Studies in Higher Education, 41*(2), 231-246.

Eldredge, N. (2012). Mathgen paper accepted. *That's Mathematics*.

Gehrmann, S., Strobelt, H., & Rush, A. M. (2019). GLTR: Statistical Detection and Visualization of Generated Text. *arXiv preprint arXiv:1906.04043*.

Hiney, M. (2015). Research integrity: what it means, why it is important and how we might protect it. *Strasbourg: Science Europe*.

Kelly-Bootle, S. (2005). Call That Gibberish. *Queue, 3*(6), 64-ff.

Labbé, C. (2010). Ike Antkare one of the great stars in the scientific firmament.

Lavoie, A., & Krishnamoorthy, M. (2010). Algorithmic detection of computer generated text. *arXiv preprint arXiv:1008.0706*.

Malviya, S., & Tiwary, U. S. (2016). Knowledge Based Summarization and Document Generation using Bayesian Network. *Procedia Computer Science, 89*, 333-340.

Mugan, J. (2018). Generating Natural-Language Text with Neural Networks. *Noteworthy - The Journal Blog*.

Nguyen-Son, H.-Q., Tieu, N.-D. T., Nguyen, H. H., Yamagishi, J., & Zen, I. E. (2017). *Identifying computer-generated text using statistical analysis.* Paper presented at the 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC).

Peterson, M. (2007). Best practices for ensuring scientific integrity and preventing misconduct. *Essays on Ethical Standards, 21*(1).

Singh, S., & Remenyi, D. (2016). Plagiarism and ghostwriting: The rise in academic misconduct. *South African Journal of Science, 112*(5-6), 1-7.

Smith, T. (2018). More states opting to'robo-grade'student essays by computer. *Retrieved from NPR: https://www. npr. org/2018/06/30/624373367/more-states-opting-to-robo-grade-student-essays-by-computer*.

Spinak, E. (2014). In the beginning it was just plagiarism – now its computer-generated fake papers as well. *SciELO in Perspective*.

Stribling, J., Krohn, M., & Aguayo, D. (2005). Scigen-an automatic cs paper generator.

Tien, N. M., & Labbé, C. (2015). SciDetectTM Documentation.

Tien, N. M., & Labbé, C. (2018). Detecting automatically generated sentences with grammatical structure similarity. *Scientometrics, 116*(2), 1247-1271.

Van Noorden, R. (2014). Publishers withdraw more than 120 gibberish papers. *Nature, 24*.

Whittaker, Z. (2019). OpenAI built a text generator so good, it's considered too dangerous

to release. *TechCrunch*.

Williams, R., & Nash, J. (2009). *Computer-Based Assessment: From Objective Tests to Automated Essay Grading. Now for Automated Essay Writing?* Paper presented at the International United Information Systems Conference.

Xiong, J., & Huang, T. (2009). *An effective method to identify machine automatically generated paper.* Paper presented at the 2009 Pacific-Asia Conference on Knowledge Engineering and Software Engineering.