This is the peer reviewed author accepted manuscript (post print) version of a published work that appeared in final form in:

Abd-Elaal, El-Sayed, Gamage, Sithara HPW, Mills, Julie E 2022 'Assisting academics to identify computer generated writing', European Journal of Engineering Education, online, pp. 1-21

This un-copyedited output may not exactly replicate the final published authoritative version for which the publisher owns copyright. It is not the copy of record. This output may be used for non-commercial purposes.

The final definitive published version (version of record) is available at:
https://doi.org/10.1080/03043797.2022.2046709

Persistent link to the UniSA Research Outputs Repository:
https://researchoutputs.unisa.edu.au/11541.2/28555

General Rights

Copyright and moral rights for the publications made accessible in the UniSA Research Outputs Repository are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the persistent link identifying the publication in the UniSA Research Outputs Repository

If you believe that this document breaches copyright, please contact us and provide details. We will remove access to the work immediately and investigate your claim.

Assisting Academics to identify Computer Generated Writing

El-Sayed Abd-Elaal[1,2], Sithara H.P.W. Gamage[1], Julie E. Mills[1]

[1]UniSA STEM, University of South Australia, Australia

[2]Structural Engineering Department, Mansoura University, Mansoura, Egypt

Corresponding Author Email: ElSayed.Abd-Elaal@unisa.edu.au

Abstract

Authentic writing is an important aspect in education and research. Unfortunately, academic misconduct occurs among students and researchers. Consequently, written articles undergo certain detection measures and most teaching and research institutions use a range of software to detect plagiarism. However, state-of-the-art Automatic Article Generator (AAG) writing powered by Artificial Intelligence (AI) provides a new platform for new types of serious academic misconduct that cannot be easily detected and even if they are detected, can be hard to prove. The main objective of this study is to raise awareness of these tools among academics. This paper first explains the features of AAG writing, then investigates whether academics can distinguish AAG writing from human writing and whether raising the awareness of AAG between academics can improve their ability to detect AAG writing. A case study showed how difficult it is for academics with no knowledge of AAGs to identify this writing. A survey was used to indicate how a training session can improve the ability of detecting AAG writing. The results show that raising awareness training increased the academics' ability to detect AAG writing. Lastly, the possible solutions to mitigate the academic integrity issues associated with AAG writing have been discussed.

1. Introduction

Jeremy Stribling and his colleagues established an automatic article generator (AAG) which cobbled 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- Cybernetics and Informatics (WMSCI) Conference on Systemics (2005). Unfortunately, those papers were accepted in this conference (Ball 2005; Nguyen Minh and Labbé 2018). Since then, significant steps have been taken to further develop article generators for useful purposes. As an example, Malviya and Tiwary (2016) developed a system using AAG to generate a research article that can summarise previous studies. Other researchers used AI techniques for developing tools to mark students' essays, which are called "robo-graders" Smith (2018).

Unfortunately, some people have continued to misuse AAG tools. Since the development of the first AAG in 2005 by Stribling, Krohn, and Aguayo (2005) to 2014, the publishers Springer and Institute of Electrical and Electronics Engineers (IEEE) have removed more than 120 papers from their subscription services after discovering that those papers were computer-generated nonsense (Oransky 2014). These kinds of papers can easily deceive inexperienced readers because they have proper sentences and structure as in a normal scientific paper . For example, Ike Antkare presented more than 100 publications in a year, bringing his name to the 21st position of the most cited scholars in the world, which is higher than Einstein (Labbé 2010). However, Antkare's ruse was later exposed and he lost his prestigious H-index and search engines removed his papers from databases. Another example is Philip Parker, who 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 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. Although these books are officially admitted for selling, Cohen (2008) reported a complaint from a purchaser who is from Perth, Australia, who said "The book is more of a template for 'generic health researching' than *anything specific to rosacea. The information is of such a generic level that a sourcebook on the next medical topic is just a search and replace away*". Philip Parker responded to this purchaser "*If you are good at the Internet, this book is useless. There are people who aren't Internet savvy who have found these guides useful*." (Cohen 2008).

Currently, there are a significant number of AAGs providing help with their assignments to students. Some examples are ("Essaybot: Free Essay Writing Tool | Essay Typer & Samples" n.d.) and MyAssignmentHelp.com (AssignmentHelpCompany 2007). There are other generators which are specialised in writing research articles, and there are even some websites e.g. Digital Marketing Tipsy ("BestArticleGeneratorSoftware" n.d.) which advertise and rank those tools.

Detecting a work generated by AAG creates an extra burden on teachers, researchers and anti-plagiarism tools. The OpenAI (an AI research company) 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).

Europe (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 plagiarism, detecting papers generated by AI needs a vigorous teacher/ reviewer.

Ghost writing (getting someone else to do the writing), also known as the 'paper mill' was one of the main issues of academic integrity (Lines 2016; Ali and Alhassan 2021). However, artificial writing using AAG tools is prospering and adding another burden to academic integrity. It may become the dominant cheating type as it is cheaper, can be produced instantly, is hard to detect and in many cases, users may not consider this as cheating.

The aim of this paper is to highlight the features of AAGs, then raise awareness among academics about AAG tools. The paper investigates the following research questions:

1. Can academic staff without knowledge of AAG distinguish the AAG produced writing from human writing?
2. Can adequate training for academics improve the ability to detect AAG writing?

2. Methods

The first research question was inadvertently answered without a planned method where a PhD student who was suspected of using an AAG initiated this research. A detailed conversation relevant to this case study is presented in the Appendix. The conversation highlights how difficult it is for academic staff without knowledge of AAG to detect such writing. Subsequently, it was identified that the knowledge of AAG among academics is very poor. Several academics were involved in the investigation detailed in the case study, including supervisors, reviewers, and academic integrity officers. They did not recognise the AAG writing until investigative work by one of the PhD supervisors raised awareness and concern. Once awareness was raised, the resultant investigation led to all supervisors agreeing that there was a nonsense writing issue.

As a result of the case study experience, this research was initiated to raise awareness of AAG among academics, as well as to train them to identify this type of writing. A detailed method was designed to answer the presented research questions in this study. A two-part survey was carried out to investigate whether training sessions can improve this knowledge.

A total of 21 academics attended the first training session, three of them did not write their unique code and therefore their answers were eliminated, meaning that the questionnaire results of 18 academics are presented. To collect more data for the research, another training session was conducted at a different campus with a different group of academic staff. The second session included 12 academics after incomplete questionnaires were eliminated. The

first session included mixed disciplines, (i.e health science, engineering, marketing and commerce), while the second session only included engineering disciplines.

For this study, a randomly chosen AAG tool, the Article Generator Pro ("Article Generator Pro: Auto Content Creation Tool"), which only provides a part of a paragraph as free of charge was used. This tool needs inputs such as keywords, word limit and reference style before it generates writing. The keywords “Artificial intelligence tool, cheating, academic integrity” were used to generate some writing and four chosen generated phrases are shown in Figure 1. For the purposes of this paper those highlighted by the letter R are the robot generated writing. Then, two phrases written by the authors related to the same topic (highlighted by the letter H for human) were added to the mix of four AAG generated sentences to create the survey form as shown in Figure 1. When the survey form was distributed to the participants, obviously the H or R was not highlighted, as the purpose was for participants to select their choice from the two options.

The training session was organised to provide the participants with information about AAGs and the academic integrity issues associated with it. The first 5 min of the training session was to introduce the academic integrity issues associated with AAG tools. Before describing the details of AAG writing, the first part of the survey was conducted. The participants were given 3 to 4 min to decide which paragraphs were robot writing and which were human writing. Then, a 15 min presentation was given to describe the concepts behind AAG and the suggested approaches to identify the generated writing. At the end of the workshop, the same survey was conducted again to see if there was any improvement in the ability to identify AAG writing. The training in between the two-part survey is detailed in the section below.

Coding

Please use this section to enter your unique code that the computer can read for statistical purposes but which ***cannot be used to identify you.***

- *Please write the day you were born, e.g. 07*
- *Please write your initials, e.g.* (EA)

One of the artificial intelligence writing tools has been used to generate some paragraphs with the keywords "Artificial intelligence tool cheating academic integrity". Can you please decide which paragraphs are written by Robot and which paragraphs are written by a human? (Put a circle around H if you think it is written by human and put a circle around R if you think it is written by robot).

"Why did we go down this path? When we analysed cheating behaviour at a macro level, we found that nearly every behaviour 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.."	***H or R***
"Ethical writing is an important matter in education and research. Unfortunately, academic cheating occurs among both undergraduate and postgraduate students. Consequently, written essays and articles undergo certain detection measures and most teaching and research institutions use a range of software to tackle plagiarism. However, the state-of -art artificial intelligence provides a new platform for new types of serious acad…."	***H or R***
"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 parameters 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..."	***H or R***
"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. 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 che..."	***H or R***
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.	***H or R***
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...	***H or R***

Figure 1: Survey form

3. Training

Information for the training was obtained from a literature review and the authors' experience in detecting AAG writing. The training included information about a) the origin of the AAG, b) examples of the positive and negative uses of AI in writing, c) how does an Automatic Article Generator (AAG) work?, d) how to identify AAG writing?, and e) possible approaches that an individual or institution should follow to deal with academic integrity issues associated with AAG tools.

The information about the origin of the AAG and positive and negative examples of the use of AI writing explained in the training were detailed in the introduction section. The rest of the information discussed in the training is detailed in section 3.1 and 3.2. Information presented in section 3.3 was briefly covered during the training because the main focus of the training session was raising awareness of AAGs.

3.1. How an Automatic Article Generator (AAG) works?

During the training, it was highlighted that AAG is an online computer software program that has access to a large database including old and new academic articles, books, reports, newspaper, web pages, etc. These computer programs search the databases according to the given keywords, or the title of the suggested research topic and then they compose the 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. Research indicates that the non-sophisticated reader could be deceived by the effective use of jargon (Kelly-Bootle 2005). These computer programs also 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. In addition, 'generating natural language techniques', which develops wording very close to human-crafted phrases, may be implemented in these tools. The information gathered from the literature was presented in the training. The summary slide used in the training is shown in Figure 2.

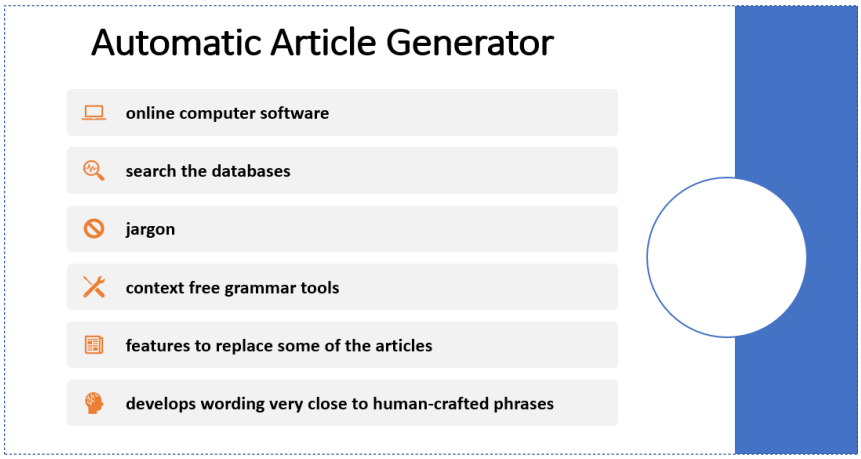


Figure 2. How an Automatic Article Generator works

During the training it was also explained that these computer programs use private techniques to summarise the collected data. These techniques might be similar or further advanced than the Malviya and Tiwary (2016) method which can be used to generate summarised research papers. This 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, introduction, methodology, and so on as shown in Figure 3.

The fully generated articles have a realistic structure, a set of references, reasonable figures and format including abstract, introduction, methodology, conclusion and reference list. Both Mathgen (Nathaniel 2012) and SCIgen (Stribling, Krohn, and Aguayo 2005) can generate full mathematics research papers and computer science research papers, respectively. Both are free tools and can generate hundreds of papers in a few hours. However, both generators are not sophisticated enough to generate articles using specific keywords, hence it seems that, both Mathgen and SCIgen don’t use lots of AI’s capabilities. During the training, three articles generated by the Mathgen tool were presented to the audiences.

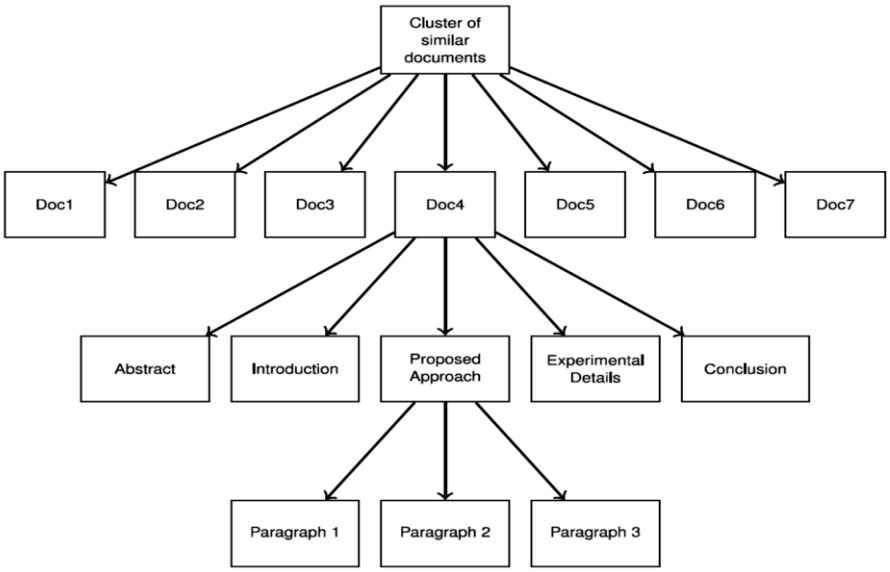


Figure 3: Structure of Knowledge Tree (Malviya and Tiwary 2016)

3.2. How to Identify Automatically Generated Writing?

Several concepts were presented in the training session to explain how to identify automatically generated writing. Firstly, the conclusion of Mugan (2018) was highlighted, namely that the computer conceptual systems are not yet rich enough to represent what a little child is aware of. Then a further reflection regarding this concept was presented, for example that 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 the phrases they collect. Even super-intelligent robots do not know when to laugh or when to cry unless specific scenarios are embedded in their memory. The text produced by AAG is not coherent on a large scale and it often does not make sense, even though it might be rich with information.

After emphasising the above concept, the authors summarised the below approaches during the training session, as they can be applied for detecting the computer generated writing. While some approaches are not viable without computer software, they were introduced to enrich the knowledge of the audiences.

3.2.1. Investigate topological properties of the writing

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, Nguyen and Labbé (2018) investigated the similarity in the grammatical structure of sentences to detect the sentences that are created using a Probabilistic Context-Free Grammar software. Although, these approaches helped to catch many gibberish papers, and in particular 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 (Nguyen and Labbé 2018). Nguyen-Son et al. (2017) reported a significant development of AAGs, in that recent AAGs can use natural language techniques which develop wording very close to human-crafted phrases. Consequently, such developments reduce the ability to detect AI writing by checking the grammatical structure of sentences.

Schuster et al. (2020) applied stylometry approaches (capturing the stylistic differences between machine-written-text and human-written text) to detect machine-generated writing, and concluded that the stylometry approche is limited against machine-generated writing.

3.2.2. Statistical Analysis of the writing

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. (2020) investigated fake online reviews and found that GLTR can accurately detect fake reviews which are developed by computer generator writing. However, in the current study, the authors tested the previously generated paragraphs. Unfortunately, the GLTR 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, the human-written articles should have keywords from the title of the cited articles and that certain keywords might be preferred by the writer throughout a manuscript. That suggestion agrees well with what is known as a writer's personal tic (tic' means some specific words that a writer tends to use more than usual, for example, "also", "significant", "important", "however", "could", "agree" and "Unfortunately"). However, these approaches may not be significant for a short piece of writing.

The authors highlighted the Lavoie and Krishnamoorthy (2010) approach (investigating patterns of keywords and overused words) during the training session because AAG cannot easily create those keyword patterns, despite developers being aware of it.

3.2.3. Investigate the references

Xiong and Huang (2009) noted that most of the references included in the papers generated by SCIgen (Stribling, Krohn, and Aguayo 2005) are fake and do not exist on the Internet. However, the subsequent versions of AAG tools have remedied this defect. Nevertheless, 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, different languages or from many internet resources). Added to this, AAG writing could include words that give different meanings based on a context, but not a clear message with a firm focus.

Also, careful investigation of 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 in fact close to the actual topic.

Rogerson (2017) revealed some clues for detecting contract cheating in student work by investigating referencing irregularities. Some of these clues agree well with AAG writing features such as irrelevant sources and the access date on the internet sources. Rogerson (2017) further indicated that reference list entries with a certain year and access date of an older year should be flagged as an irregularity. However, in the case of AAG writing, all access might have happened in a day, while it is expected that it would take weeks for such tasks undertaken by a human researcher. Lines (2016) investigated ghost writing and identified other features such as the use of American English in essays submitted in an Australian university. This feature has been noted in Figure 1 and is applicable to identify artificial writing, unless the developer or the students consider it and make manual changes.

In the training session, the above information was summarised while emphasising that the AAG writing can be difficult to recognise in review type of work, such as a literature review. In the case of experimental or numerical work, supervisors need to assess the raw data of the reported findings, particularly if AAG writing is detected or suspected.

3.3. Possible approaches to deal with academic integrity issues associated with AAG tools

In the past decade, paraphrasing tools and language translation tools imposed burdens on academic integrity (Rogerson and McCarthy 2017; Prentice and Kinden 2018). Now, AAG tools are adding extra burdens. Google trends point out the 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 as stated below. These were briefly mentioned in the training sessions.

3.3.1. Importance of Training educators

It is important to raise awareness among educators about these tools because many educators may not be aware of them. It is found that markers can often detect contract cheating when asked to do so, and training may be helpful in improving their accuracy of detecting the cheating (Dawson and Sutherland-Smith 2018; Curtis et al. 2021). Williams and Nash (2009) suggested that markers' suspicions are a crucial factor in addressing contract cheating. Therefore, raising the awareness of AAG and training the markers to identify such articles are very important aspects.

One way of raising awareness is to hold workshops for educators to provide a clear understanding of how AAG works and to explain the characteristics of the AI-generated writing. Then if a marker notices any features of artificial writing, such as a significant number of references that a student has included but it is not possible for the student to actually gather and read all those references in the timeframe, it will raise suspicion. Also, if a student's work has flawless grammar and is formatted incredibly well while the fundamental ideas are not there or do not make sense, then it should raise alarm.

3.3.2. Minimising opportunities for cheating

Educators need to consider the underlying reasons for plagiarism and should address these by providing extra support. The primary approach to address reasons for plagiarism, in the authors' opinion, is a more humanitarian approach. This may require several personal and institutional changes. Bretag et al. (2018) and Harper et al. (2019) conducted two large contract cheating surveys for Australian university students and staff, respectively. They concluded that core standards of academic integrity are: the development of teaching and learning environments that bring strong student-teacher relationships, reduce opportunities for cheating through curricular and assessment design, and address the well-recognised language and learning needs of non-native English speakers.

Another study has highlighted five conditions that cause academic dishonesty. They are academic aptitude, perseverance, personal issues, competing objectives, and self-discipline (Amigud and Lancaster 2019). Some scholars suggest that learners who come from certain countries have a more accepting and permissive attitude towards cheating at universities (Ehrich et al. 2016); (Duff, Rogers, and Harris 2006). Another study indicates that students cheat because of 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.

It is also important to understand whether the students realise that they are committing an offence of academic misconduct. Technology and Internet resources are factors that are linked to study habits (Minichiello et al. 2019) and using readily available resources may not seem like misconduct for some learners. A recent study found that a large percentage of students who study face-to-face courses and online courses do not consider their behaviours to be a violation of academic misconduct (Burgason, Sefiha, and 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. Carpenter et al. (2006) concluded that the academic integrity policy must be revealed and communicated clearly to the students in the classroom. However, conducting a direct discussion with students regarding AAG tools may not be a good approach, as some students may not know of the existence of these tools and such a discussion could be an unintentional advertisement of these tools. It might be better to decide whether to discuss this type of cheating with students or not depending on the existence of some cases of this type of cheating. If there are no or limited cases, then there is no need to discuss this type of cheating with the students.

A study conducted to investigate perceptions and motivations for cheating suggested that the general apathy students may have towards their studies causes cheating and online classes increased this 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.

Giving students less opportunity to cheat is another part of the solution. This can be achieved by structuring the assessment types that are less likely to result in cheating. Bretag et al. (2019) found that assessment types such as review articles, literature analysis, integrated knowledge/skills, real world tasks and no right answer tasks are more likely to be outsourced by students. Ellis et al. (2020) investigated the relationship between authentic assessment and academic integrity and concluded that, students are routinely outsourcing assessment tasks regardless of the authenticity requirements. However, the authors believe that there is a relationship between the technique of cheating and assignment type. For example, if the student is looking for someone else to do the assessment, the type of assignment is irrelevant. However, in the case of using AAG tools, the assessment type is an important parameter.

There are types of assessment that cannot be easily produced by article generators. A series of small graded tasks, in-class activities, personalised assignments, customised practicals, interviews/vivas, assignments based on active learning methods (Heylen and Vander Sloten 2013), assignments where students have to provide argumentative reasoning (Leite et al.

2011), and assignments linked to technical and social competencies are less likely to result in cheating. Therefore, enhancing the curriculum with assessments which are difficult to cheat can be a solution to AAG tools, as well as some other forms of academic misconduct. In addition, creating an environment to motivate students is a good option. Krou, Fong, and Hoff (2021) concluded that academic integrity is positively related to motivation and extrinsic goal orientation.

Finally, instigating severe penalties for offenders can also reduce the number of academic misconduct cases. Chirikov, Shmeleva, and Loyalka (2020) concluded that severe punishments greatly improve students' attitudes toward academic dishonesty. However, this should not be the primary focus of educators, as this will not address the reasons why people use these unauthorised services and commit academic misconduct. Also, it might increase the education administration work.

3.3.3. Developing tools for detecting AAG writing

Altheneyan and Menai (2020) conducted an evaluation of paraphrasing identification techniques. They concluded that further research would represent a promising research direction for both paraphrasing identification and automatic plagiarism detection. Paraphrasing identification means determining if two text segments have the same meaning. It's one of the most interesting and fascinating challenges in the field of natural language processing (NLP) (Shahmohammadi, Dezfoulian, and Mansoorizadeh 2021). The solution for this problem will provide tools to identify different types of cheating such as text summarisation, machine translation, and question answering. It is worth developing tools to identify the characteristics of automatically generated writing and introducing more processes to flag them for the markers.

Such tools will save the time of markers who are already under pressure. However, these tools may not give a conclusive decision, therefore, educators have to be vigilant and judges are needed for investigating any suspicious writing.

3.3.4. Academic misconduct policies of Universities

A brief review of Australian university policies indicated that few institutions currently have a warning for using artificial writing tools in their academic misconduct policies. One positive example is at Flinders University, South Australia which includes the following wording in their contract cheating clause "…*submitting an assessment task generated by an algorithm, computer generator or other artificial intelligence…"* (Flinders-University 2018). However, other universities do not have a clause against AI witting at all or have clauses that may not be adequate enough, for example, the Sydney University policy which warns only against

ghost writing, says that "…*Contract cheating involves getting* someone else to complete part or all of *your work and then submitting the work as if you had completed it yourself…*" (USYD 2019).

Another example is, the University of South Australia policy, which warns against ghost writing and falsifying documents or data but does not refer to AAG tools, saying "*Academic misconduct occurs when staff or students disregard the values of academic integrity…. Some examples include plagiarism, cheating in exams, copying or falsifying documents or data, getting other people to write your work or doing someone else's work for them*" (UniSA 2021).

Furthermore, the declaration that all candidates make in their thesis does not mention using AI tools. UniSA's declaration says: *"I declare that: this thesis presents work carried out by myself and does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; to the best of my knowledge it does not contain any materials previously published or written by another person except where due reference is made in the text; and all substantive contributions by others to the work presented, including jointly authored publications, is clearly acknowledged." (UniSA 2021)*

Therefore, Academic integrity policies need to include a warning against the use of *artificial intelligence writing.* However, revising Academic Integrity Policies may not be enough, as students are likely to deny their action, as strong conclusive evidence is difficult to gather. Therefore, the following process is suggested:

1. Updating Academic Integrity Policies at universities to include a warning in relation to the use of artificial intelligence tools for automatically generating writing.
2. Preparing a training module and workshops for training tutors, markers and course coordinators, so that they will be aware of this kind of misconduct.
3. Establishing a review panel from qualified academics. This review panel should be fully aware of article generator tools. The panel members should pass an intensive series of AAG training sessions.
4. If a tutor or marker observes a cheating issue and the course coordinator agrees to investigate this request then, the course coordinator should send a request to the Academic Integrity Officer.
5. The Academic Integrity Officer should forward the request to the review panel for investigation. If the majority of the review panel agree that the writing is artificially generated then a penalty will apply that is the same as contract cheating penalties.
6. The student has the right to appeal and in this case, the appeal process should require the student to run a substantial presentation (say 30 minutes) explaining the ideas, concepts, critical thinking and the conclusion of the writing under investigation. This should be done in front of the review panel, program director, course coordinator and

the marker and they should discuss the writing issues with the student. The authors believe that, if the writing is synthesised by artificial tools, the student will not be able to explain the meaning of the writing and in this case, the Academic Integrity Officer may recommend a suitable penalty.

4. Analysis of robot writing in questionnaire form

It is also worth mentioning that the generated paragraphs in questionnaire form are not entirely robot writing. The robot chooses pieces of existing human texts and then copies, paraphrases and pastes them together to generate sentences within given conditions.

Article and essay generator tools have various features or conditions so users can get the best quality they aim to, for example, Article Generator Pro ("Article Generator Pro: Auto Content Creation Tool") has multiple features as shown in Figure 4. These features mainly include : a) Research Job feature which gives an option to decide how deep the research should be, b) Text Rewrite/Article Spinning feature which replaces the original article with random synonyms, a higher ratio means more words will be replaced (must be used to pass plagiarism), b) Reorder Sentence feature which rearranges sentences but the article may no longer make sense, and d) Relevancy Mode feature which generates an article that matches the best to the searched topic.

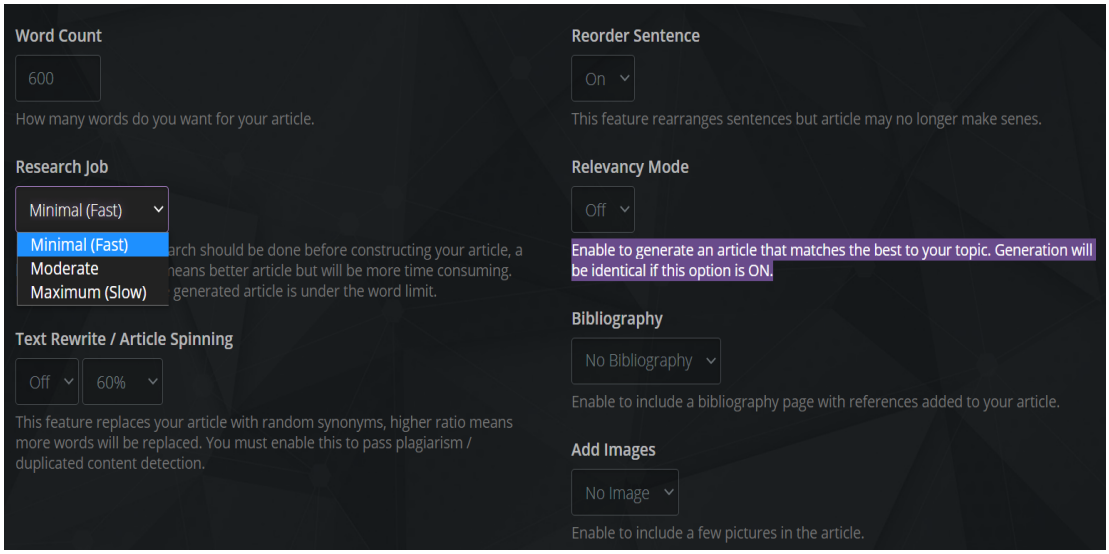


Figure 4: Article generator features ("Article Generator Pro: Auto Content Creation Tool")

The presented paragraphs in the questionnaire form were produced with minimum use of replacing features so it will be easier to revisit the original resources via deduction plagiarism tools. However, it might be difficult for educators to determine the act of the robot in such paragraphs during the training session because of the minimal changes in the original writing. Figure 5 shows a comparison between the original text and the text developed by the AAG.

AAG tools were able to generate texts that best match the topic. However, there are some defects in their products even when choosing the minimal robot actions from the original manuscript as shown below:

- Text 1, AAG tool replaced some words with random synonyms. However, the used synonyms like "*manner"* and "*solution"* do not match the concept of the paragraph.
- Text 2, AAG tool compiled two texts together from two different sources and there is an incongruous shift between the two parts.
- Text 3 seems fine, even though five words were replaced with their synonyms
- Text 4, AAG tool compiled three texts together from two different sources and there is an incongruous shift between the last two phrases.

1. Original text	*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* ***recognize*** *those specific actions. Since we can train a human to recognize these patterns of* ***behavior****, we knew we could train a machine to do the same. However, in our live proctoring* ***solution****, technology is* ***used*** *to enhance human* ***abilities.*** *Source: (ProctorU 2019).*
1. AAG text	*"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.."*
2. Original text	I. *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. ~~As such, the challenge of spotting AI-generated papers is very similar to the challenge of spotting human-ghostwritten papers.~~ Source: (Bailey 2019)* II. *When we look back on writing papers for school, at some point, we have all experienced the feeling that the work is non-transferable. ~~How would that deep dive into The Lovesong of J. Alfred Prufrock prepare us for life? The answer: picking apart a line of poetry—or engaging in any other form of academic writing—teaches us to think critically and translate those thoughts into words on paper.~~ Source: (Turnitin 2017)*
2. AAG text	*"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..."***
3. Original text	*Though the idea of robots writing* ***school*** *papers might seem to be the* ***realm*** *of science fiction, the truth is robots are already writing content. In September 2017, the Washington Post announced their AI reporter, dubbed Heliograf, had* ***penned*** *some 850 stories in the* ***prior*** *year. This* ***included*** *some 300 reports from the Rio Olympics. Source: (PG 2019)*
3. AAG text	*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.*

4. Original text	*I. Even top tertiary institutions like Harvard, Stanford University, and Duke University among others have not escaped cheating scandals. ~~Several studies point to a similar problem in high schools. Schools have been looking into ways to prevent high-tech cheating before it happens, and now artificial intelligence can also be leveraged for this purpose. There is a number of new technologies that universities can leverage to prevent cheating and plagiarism~~. Source:(Lynch 2019)* *II. ~~Cheating is a serious problem in American high schools and universities.~~ 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. ~~Even top tertiary institutions like Harvard, Stanford University, and Duke University among others have not escaped cheating scandals. Several studies point to a similar problem in high schools~~. Source: (Schroeder 2019)* *III. Turnitin's various products are in wide use: more than 30 million students from more than 15,000 K-12 and higher education institutions in 153 countries. ~~Its network of clients includes educational and research institutions, corporations and governments. The web service has a phenomenal database of literally millions of documents, essays, research papers, academic articles and surveys, web pages and more to compare any writing samples for suspicious correspondence. This database also includes student papers submitted on Turnitin, which run into millions~~. (Lynch 2019)*
4. AAG text	***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 5: A comparison between the original text and corresponding developed AAG text

5. Results and discussion

The presented case study and survey revealed the academic's ability to detect AAG writing and showed the importance of training sessions. Detailed results and discussion regarding the case study and the two research questions are provided in the following sections.

5.1. Case study results

The case study showed that the supervisors and the reviewers of the student's proposal and publication were not able to detect the AAG writing until they gained awareness of AAGs at a later stage. It was notable that, even when there was agreement between the supervising panel at a later stage that the student was using AAG's tools, they didn't raise an academic integrity case because of the difficulty of providing solid evidence for this type of cheating. It was also clear in this case that AAG can result in a misuse of an institution's funds, lead to a possible loss of the supervisor's/scholar's reputation and potentially allow a fraudulent graduate to enter academic society.

- The case study provides key clues that are useful to identify AAG-produced writing. They are: Sentences or phrases that seem OK as individual items but when presented as a whole, do not make sense and lack cohesion.

- There are abrupt changes in ideas that are not connected by any logical argument. This indicated the importance of reviewing a chapter/article as a whole rather than section by section.
- References that are not relevant but contained some keywords in the title, that are too old and obscure, or that are in a language not spoken by the student. Also, too many references provided, such that a student couldn't possibly review all of them in the timeframe that they took to produce the article.
- Extensive new chapters/articles being produced in a timeframe that is much quicker than normal.

5.2. Can academic staff without knowledge of AAG distinguish the AAG produced writing from human writing?

The case study and the first part of the survey appear to provide different answers for the first research question. The case study showed that the answer to the first research question is no or limited, unless there is awareness of AAG. This answer aligns well with a recent experiment study ("What Grades Can AI Get in College?" 2020) where a panel of professionals were hired to grade samples of assignments including AAG's writing and none detected the AAG writing. On the other hand, the results of the first part of the survey showed that, the participants from the two groups have achieved some capability of 56.5% and 62.5% respectively in detecting AAG's writing, as shown in Figure 6.

However, it is important to emphasise that the results from the first part of the survey do not reflect the usual ability of participants to detect the AAG's. The trainer raised a question at the start of the sessions about whether any of the audience knew about AAG tools and none admitted having such knowledge. Furthermore, the survey form disclosed that there were examples of both AAG writing and human writing on the form. So, the participants' correct answers in the first part of the survey were expected to be influenced by the awareness-raising undertaken plus the 5 mins of the presentation before running the first survey. This indicated that even raising awareness with a short introduction was already very important. This also confirms that markers can often detect contract cheating when asked to do so (Dawson and Sutherland-Smith 2018).

5.3. Can adequate training for academics improve the ability to detect AAG writing?

It can be concluded from the two-part survey that, the participators gained knowledge regarding the AAG's writing. The average percentage of correct answers for all workshop participants in the first and second survey was 59.5%, rising to 71.1 %, respectively. The results by workshop group presented in Figure 6 indicate that the second group had a better ability to detect the robot writing both before and after the training sessions. This might be

because the second group and the presenter were all from the same discipline area, namely engineering.

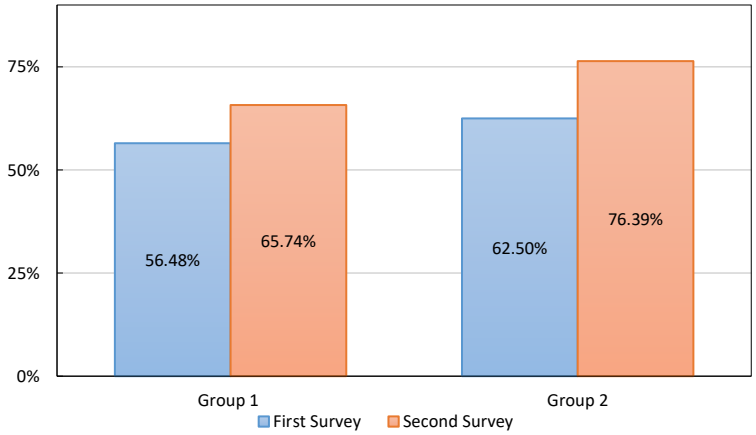


Figure 6. Percentage of correct answers for each group of participants before and after the training session

The detailed results of the two-part survey for participants in group 1 and group 2 are shown in Figure 7.a and 7.b, respectively. Comparing results before and after the training session for all participants using the Wilcoxon signed rank test returned the test score of P value = 0.11. This indicates improvement in the ability for detecting AAG's writing after the training, but it is not statistically significant.

Accordingly, the answer to the second research question- "Can adequate training for academics improve the ability to detect AAG writing?" is yes, on the basis of the study in this paper, but more data would be required to provide a more definitive answer. However, it is likely that greater improvement in detection ability would require a more advanced training program or a series of training sessions. Furthermore, the presented questions in the survey have limited robot touch – as discussed in section 4. They were produced with the minimum use of replacing features because the authors wanted to revisit the original resources via deduction plagiarism tools, which made the survey's questions challenging. So, perhaps a range of questions with different difficulty levels should be adopted and thus the improvement after training will be measured correctly. The case study also demonstrates that the ability to detect AAG writing in a more extensive piece of work such as a thesis, requires different awareness than individual paragraphs as used in the survey study.

Figure 8 shows the percentage of correct answers for each question from the two sessions. The participants demonstrated a better ability to recognise the robot writing in Q1 and Q5, while they showed less ability to recognise robot writing in Q3 and Q6. As shown in section 4, the characteristic of Q1 and Q5 is that they contain random synonym word replacements while Q3 and Q6 are characteristic of compiling a few sentences from different sources. It appears that the participants have a higher ability to detect the robot writing when AAG uses

replacement synonyms rather than combining the sentences from different sources. Therefore, further focus on detecting generated text with combining sentences is required in future training. In view of that, at the end of the first training session, some participants raised concerns regarding the paragraphs being too short and suggested that long pieces of writing were needed to easily recognise the AAG generated writing. This concern is where the participants showed a higher ability to recognise the robot texts with replacement synonyms rather than the combined texts from different sources. In the case of short piece of writing, there will be single abrupt shift between the ideas while a long piece of writing will have multiple shifts between the ideas.

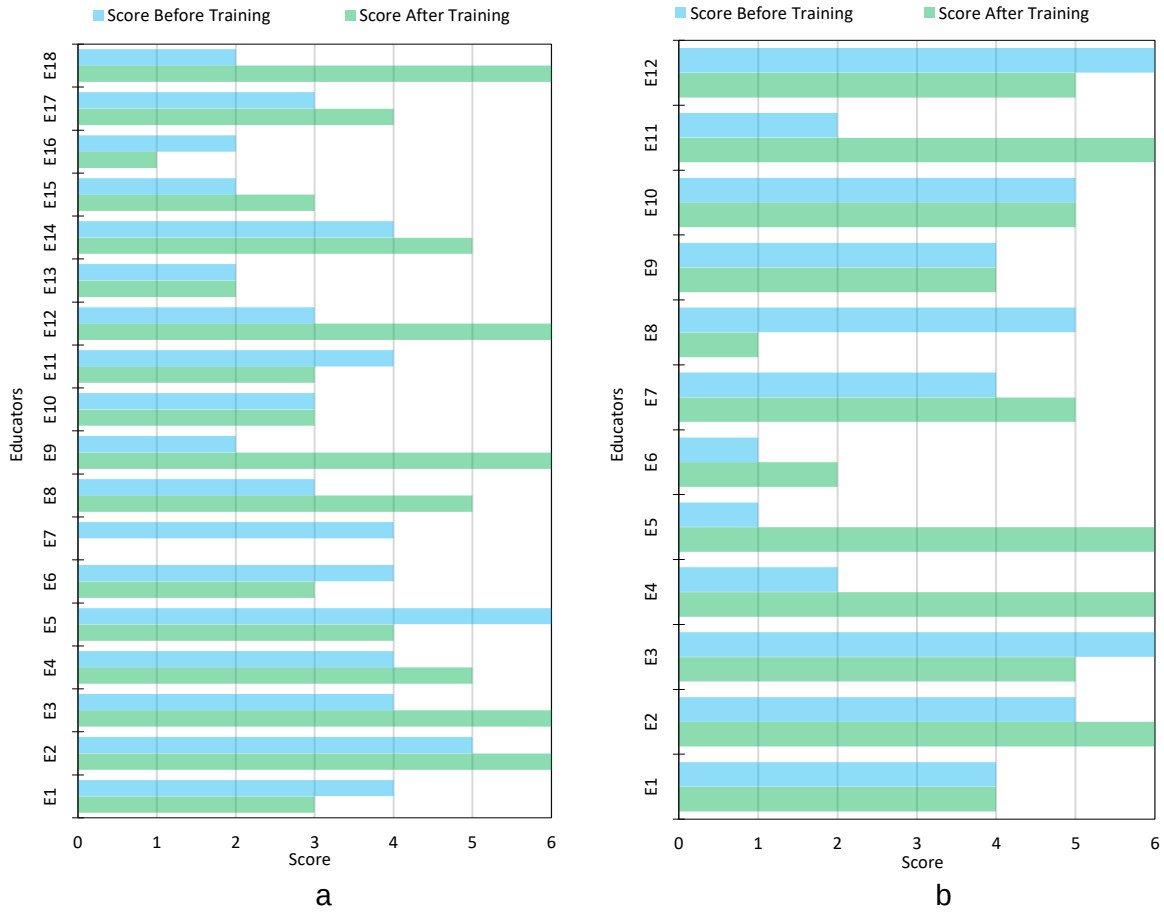


Figure 7. Academic's survey results before and after the training session a. Group 1 results, b. Group 2 results

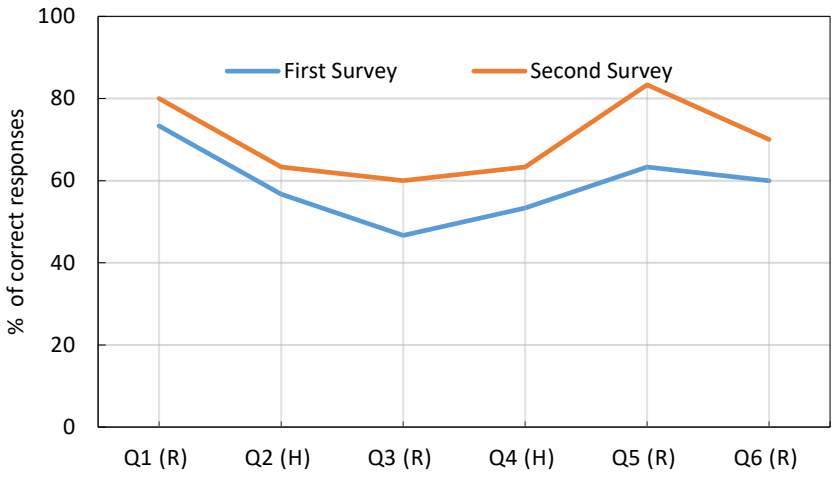


Figure 8. Percentage of correct answers for each question before and after the training session for all participants (two groups)

5.4. Study limitation and recommendations for future work

There are some limitations in this study. Firstly, the duration of the training session was short. The effective duration of training was 15 mins while the total duration of the session was 30 mins. With more time for training, more materials as detailed in this paper could have been explained to the participants. Also, section 4 was not included in the training session, but the authors recommend including it in future training. The authors recommend further studies particularly focus on section 4 (analysis of robot writing). This analysis could help academics to become familiar with robot writing styles and their faults, providing them with a better chance of recognising the robot writing.

Secondly, the introduction at the start of the training session and the survey form asked the participants to indicate which paragraphs were robot written. The introduction and this question directed the audience's attention to the fact that there was robot generated writing in the given survey form, as did the title of the training session. The authors suggest that a better approach would be to change the title of the session and the survey to a more generic one such as “How confident are you in detecting fake writing?” (or something similar) and to ask the participants to assess the paragraphs as original writing or fake writing, before mentioning anything about the AAG tools. The second survey would then be completed after the training and at this point, the terms original and fake could also be explained as equivalent to human and robot. In this case, the academics’ ability will be evaluated without the provision of initial clues, and the impact of the training session will be more effectively evaluated.

Thirdly, the presented writing in the questionnaire form should be longer and potentially include different types of robot writing (different fractions of replacing features) and from different AAG tools.

The final limitation of this study is that it was based on only two sessions, and so of course it is recommended that additional sessions be run with the modifications suggested above in order to provide further results.

6. Conclusion

This paper discusses the possible academic integrity issues associated with AAGs and possible solutions to those issues. It highlights the features of AAGs driven by AI and the techniques behind them. A case study is presented to demonstrate a situation where academics didn't have any knowledge of AAGs powered by AI and the problems this caused. Then a training workshop for educators is presented and evaluated. Finally, the responsibilities and potential techniques that could be used to mitigate the risk of possible academic misconduct associated with AI tools are discussed.

The study concludes that:

- AAGs are available to anyone and are rapidly improving the quality of their outputs. Although AAGs are powerful enough to collect the related critical information more quickly than a good researcher, these tools are not aware of the meaning of the collected information.
- There are some automatic approaches for detecting such fake writing, however, there is a need for vigorous involvement of academic staff.
- Academics should be made aware of AAG products and training workshops should be conducted to improve their abilities to detect fake writing.
- A single training session is confirmed to be a useful approach for raising awareness of AAG between academics plus improving the ability to detect AAG writing.
- The authors recommend two different types of training program. A single training session for all general staff similar to the conducted one in this study and a serious of advanced training sessions for Academic Integrity Officers, that consider the suggested recommendations for training marterials discussed in this paper.
- The teaching and research institutions should improve their plagiarism and fabrication policies against AI tools.

7. Ethics approval

The protocol of this study was reviewed and granted ethics approval and meets the requirements of the National Statement on Ethical Conduct in Human Research from the University of South Australia (Application ID 202931: Artificial intelligence and academic cheating).

References

Abd-Elaal, El-Sayed, Sithara HPW Gamage, and Julie E. Mills. "Artificial intelligence is a tool for cheating academic integrity." In 30th Annual Conference for the Australasian Association for Engineering Education (AAEE 2019): Educators Becoming Agents of Change: Innovate, Integrate, Motivate, pp. 397-403. Brisbane, Queensland: Engineers Australia, 2019.

Adelani, David Ifeoluwa, Haotian Mai, Fuming Fang, Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. 2020. Generating Sentiment-Preserving Fake Online Reviews Using Neural Language Models and Their Human- and Machine-Based Detection. Paper presented at the International Conference on Advanced Information Networking and Applications, Cham.

Ali, Holi Ibrahim Holi, and Awad Alhassan. 2021. "Fighting contract cheating and ghostwriting in Higher Education: Moving towards a multidimensional approach." Cogent Education 8 (1):1885837.

Altheneyan, Alaa, and Mohamed El Bachir Menai. 2020. "Evaluation of state-of-the-art paraphrase identification and its application to automatic plagiarism detection." International Journal of Pattern Recognition 34 (04):2053004.

Amancio, Diego Raphael. 2015. "Comparing the topological properties of real and artificially generated scientific manuscripts." Scientometrics 105 (3):1763-79.

Amigud, Alexander, and Thomas Lancaster. 2019. "246 reasons to cheat: An analysis of students' reasons for seeking to outsource academic work." Computers & Education 134:98-107.

"Article Generator Pro: Auto Content Creation Tool." 2019. Accessed 09/08/2019. https://www.articlegeneratorpro.com/.

AssignmentHelpCompany. 2021. "my assignment help." Accessed 20/06/2021. https://myassignmenthelp.com/essay-typer.html.

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

Bailey, Jonathan. 2021. "How AI Will Change Authorship and Plagiarism - Plagiarism Today." Plagiarism Today, Accessed 04/07/2021. https://www.plagiarismtoday.com/2019/02/12/how-ai-will-change-plagiarism/.

Ball, Philip. 2005. "Computer conference welcomes gobbledegook paper." Nature 434 (7036):946-.

"BestArticleGeneratorSoftware." 2021. Accessed 20/06/2021. https://digitalmarketingtipsy.com/tools/best-article-generator-software/.

Bretag, Tracey, Rowena Harper, Michael Burton, Cath Ellis, Phil Newton, Pearl Rozenberg, Sonia Saddiqui, and Karen Haeringen. 2018. "Contract cheating: a survey of Australian university students." Studies in Higher Education 44:1-20.

Bretag, Tracey, Rowena Harper, Michael Burton, Cath Ellis, Philip Newton, Karen van Haeringen, Sonia Saddiqui, and Pearl Rozenberg. 2019. "Contract cheating and assessment design: exploring the relationship." Assessment & Evaluation in Higher Education 44 (5):676-91.

Burgason, Kyle A, Ophir Sefiha, and Lisa Briggs. 2019. "Cheating is in the Eye of the Beholder: an Evolving Understanding of Academic Misconduct." Innovative Higher Education 44 (3):203-18.

Carpenter, Donald D, Trevor S Harding, Cynthia J Finelli, Susan M Montgomery, and Honor Passow. 2006. "Engineering students' perceptions of and attitudes towards cheating." Journal of Engineering Education 95 (3):181-94.

Chirikov, Igor, Evgeniia Shmeleva, and Prashant Loyalka. 2020. "The role of faculty in reducing academic dishonesty among engineering students." Studies in Higher Education 45 (12):2464-80.

Cohen, Noam. 2019. "He wrote 200,000 books (but computers did some of the work)." Accessed 05/07/2021. https://www.nytimes.com/2008/04/14/business/media/14link.html#:~:text=Philip%20M.,he%20makes%20money%20doing%20it.

Costley, Jamie. 2019. "Student Perceptions of Academic Dishonesty at a Cyber-University in South Korea." Journal of Academic Ethics 17 (2):205-17.

Curtis, Guy J., Christine Slade, Tracey Bretag, and Margot McNeill. 2021. "Developing and evaluating nationwide expert-delivered academic integrity workshops for the higher education sector in Australia." Higher Education Research & Development:1-16. doi: 10.1080/07294360.2021.1872057.

Dawson, Phillip, and Wendy Sutherland-Smith. 2018. "Can markers detect contract cheating? Results from a pilot study." Assessment Evaluation in Higher Education 43 (2):286-93.

Duff, Andrea H, Derek P Rogers, and Michael B Harris. 2006. "International engineering students—avoiding plagiarism through understanding the Western academic context of scholarship." European journal of engineering education 31 (6):673-81.

Ehrich, John, Steven J Howard, Congjun Mu, and Sahar Bokosmaty. 2016. "A comparison of Chinese and Australian university students' attitudes towards plagiarism." Studies in Higher Education 41 (2):231-46.

Eldredge, Nate. 2019. "Mathgen." Accessed 05/08/2021. https://thatsmathematics.com/blog/mathgen/.

Ellis, Cath, Karen van Haeringen, Rowena Harper, Tracey Bretag, Ian Zucker, Scott McBride, Pearl Rozenberg, Phil Newton, and Sonia Saddiqui. 2020. "Does authentic assessment assure academic integrity? Evidence from contract cheating data." Higher Education Research Development 39 (3):454-69.

"Essaybot: Free Essay Writing Tool | Essay Typer & Samples." 2021. M.essaybot.com, Accessed 20/06/2021. https://www.essaybot.com/.

Europe, Science. 2015. "Research integrity: what it means, why it is important and how we might protect it." Science Europe, Accessed 11/12/2019.

Flinders-University. 2019. "Student Academic Integrity Policy." Accessed 08/09/2019. https://www.flinders.edu.au/content/dam/documents/staff/policies/academic-students/student-academic-integrity-policy.pdf.

Gehrmann, Sebastian, Hendrik Strobelt, and Alexander M Rush. 2019. "GLTR: Statistical Detection and Visualization of Generated Text." arXiv preprint arXiv:1906.04043.

Harper, Rowena, Tracy Bretag, Cath Ellis, Philip Newton, Pearl Rozenberg, Sonia Saddiqui, and Karen van Haeringen. 2019. "Contract cheating: a survey of Australian university staff." Studies in Higher Education 44 (11):1857-73.

Heylen, Christel, and Jos Vander Sloten. 2013. "A technical writing programme implemented in a first-year engineering course at KU Leuven." European journal of engineering education 38 (6):595-607.

Kelly-Bootle, Stanley. 2005. "Call That Gibberish." Queue 3 (6):62-4.

Krou, Megan R., Carlton J. Fong, and Meagan A. Hoff. 2021. "Achievement Motivation and Academic Dishonesty: A Meta-Analytic Investigation." Educational Psychology Review 33 (2):427-58.

Labbé, Cyril. 2010. "Ike Antkare one of the great stars in the scientific firmament." In.: LIG.

Lavoie, Allen, and Mukkai Krishnamoorthy. 2010. "Algorithmic detection of computer generated text." arXiv preprint arXiv:1008.0706.

Leite, Carlinda, Ana Mouraz, Rui Trindade, José M Martins Ferreira, Ana Faustino, and Jaime E Villate. 2011. "A place for arguing in engineering education: a study on students' assessments." European journal of engineering education 36 (6):607-16.

Lines, Lisa. 2016. "Ghostwriters guaranteeing grades? The quality of online ghostwriting services available to tertiary students in Australia." Teaching in Higher Education 21 (8):889-914.

Lynch, Matthew. 2021. "How Schools Can Fight Cheating with Artificial Intelligence - The Tech Edvocate." The Tech Edvocate, Accessed 04/07/2021. https://www.thetechedvocate.org/how-schools-can-fight-cheating-with-artificial-intelligence/.

Malviya, Shrikant, and Uma Shanker Tiwary. 2016. "Knowledge Based Summarization and Document Generation using Bayesian Network." Procedia Computer Science 89:333-40.

Minichiello, Angela, Sherry Marx, Laurie McNeill, and Christine Hailey. 2019. "Exploring student study behaviours in engineering: how undergraduates prepared textbook problems for online submission." European journal of engineering education 44 (1-2):253-70.

Mugan, Jonathan. 2019. "Generating Natural-Language Text with Neural Networks." Accessed 25/08/2019. https://blog.usejournal.com/generating-natural-language-text-with-neural-networks-e983bb48caad.

Nguyen-Son, Hoang-Quoc, Ngoc-Dung T Tieu, Huy H Nguyen, Junichi Yamagishi, and Isao Echi Zen. 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), Malaysia.

Nguyen Minh, Tien, and Cyril Labbé. 2018. "Detecting automatically generated sentences with grammatical structure similarity." Scientometrics 116 (2):1247-71.

Oransky, Ivan. 21. "Unfortunately, scientific publishing is not immune to fraud and mistakes": Springer responds to fake papers story." Retraction Watch, Accessed 29/06/2021. https://retractionwatch.com/2014/02/27/unfortunately-scientific-publishing-is-not-immune-to-fraud-and-mistakes-springer-responds-to-fake-papers-story/.

Peterson, MJ. 2007. "Best practices for ensuring scientific integrity and preventing misconduct." In Essays on Ethical Standards.

PG. 2021. "How AI Will Change Authorship and Plagiarism – The Passive Voice." Thepassivevoice.com, Accessed 04/07/2021. https://www.thepassivevoice.com/how-ai-will-change-authorship-and-plagiarism/.

Prentice, Felicity M, and Clare E Kinden. 2018. "Paraphrasing tools, language translation tools and plagiarism: an exploratory study." International Journal for Educational Integrity 14 (1):1-16.

ProctorU. 2021. "How Technology is Taking Education & Proctoring to New Heights." OLC, Accessed 04/07/2021. https://onlinelearningconsortium.org/how-technology-taking-education-proctoring-new-heights/.

Rogerson, Ann M. 2017. "Detecting contract cheating in essay and report submissions: process, patterns, clues and conversations." International Journal for Educational Integrity 13 (1):10.

Rogerson, Ann M, and Grace McCarthy. 2017. "Using Internet based paraphrasing tools: Original work, patchwriting or facilitated plagiarism?" International Journal for Educational Integrity 13 (1):2.

Schroeder, Ray. "How Schools Can Fight Cheating with Artificial Intelligence - Matthew Lynch, Tech Edvocate." Continuingedupdate.blogspot.com, Accessed 04/07/2021. https://continuingedupdate.blogspot.com/2019/08/how-schools-can-fight-cheating-with.html.

Schuster, Tal, Roei Schuster, Darsh J Shah, and Regina Barzilay. 2020. "The limitations of stylometry for detecting machine-generated fake news." Computational Linguistics 46 (2):499-510.

Shahmohammadi, Hassan, MirHossein Dezfoulian, and Muharram Mansoorizadeh. 2021. "Paraphrase detection using LSTM networks and handcrafted features." Multimedia Tools Applications 80 (4):6479-92.

Smith, Tovia. "More States Opting To 'Robo-Grade' Student Essays By Computer." Npr.org, Accessed 05/07/2021. https://www.npr.org/2018/06/30/624373367/more-states-opting-to-robo-grade-student-essays-by-computer#:~:text=More%20States%20Opting%20To%20'Robo,Student%20Essays%20By%20Computer%20%3A%20NPR&text=%23NowPlaying-,More%20States%20Opting%20To%20'Robo%2DGrade'%20Student%20Essays%20By,certainly%20handle%20grading%20students'%20essays.

Spinak, Ernesto. 2019. "In the beginning it was just plagiarism – now its computer-generated fake papers as well." SciELO in Perspective, Accessed 01/08/2019. https://blog.scielo.org/en/2014/03/31/in-the-beginning-it-was-just-plagiarism-now-its-computer-generated-fake-papers-as-well/#.YOKn5OgzY2w.

Stribling, Jeremy, Max Krohn, and Dan Aguayo. "Scigen-an automatic cs paper generator." Accessed 05/07/2021. https://pdos.csail.mit.edu/archive/scigen/#about.

Turnitin. 2021. "Podcast: The Written Word." Accessed 04/07/2021. https://www.k12dive.com/spons/podcast-the-written-word/503105/.

UniSA. 2019. "Academic Integrity Module." Accessed 01/10/2019. https://lo.unisa.edu.au/mod/book/view.php?id=252142&chapterid=39188.

USYD. 2019. "Academic dishonesty and plagiarism." Accessed 01/10/2019. https://sydney.edu.au/students/academic-dishonesty/contract-cheating.html.

"What Grades Can AI Get in College?". Education Reference Desk, Accessed 08/12/2021. https://www.eduref.net/features/what-grades-can-ai-get-in-college/.

Whittaker, Zack. 2019. "OpenAI built a text generator so good, it's considered too dangerous

to release." TechCrunch, Accessed 23/08/2019. https://techcrunch.com/2019/02/17/openai-text-generator-dangerous/.

Williams, Robert, and John Nash. 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, Sydney, Australia.

Xiong, Jiping, and Tao Huang. 2009. An effective method to identify machine automatically generated paper. Paper presented at the 2009 Pacific-Asia Conference on Knowledge Engineering and Software Engineering, Shenzhen, China.

Appendix

This section presents a case study that takes the form of a "dialogue" between the first author and a senior academic. It discusses a suspected case of academic misconduct with automatically generated articles being submitted by a PhD student. The dialogue explains possible techniques that the student might follow to cover up/back up AAG tools. The author “A” conducted several discussions with a senior academic “S” about the process they went through with the student and what issues aroused suspicion about the use of AAG tools in writing the thesis. These have been summarised as a single ‘conversation’ that follows. It should be noted that the PhD student in this case was not a native English speaker, and neither were three of the four members of the supervisory panel. Hence some allowance was made during the supervision for the student to develop the kind of English writing skills needed for a complex piece of writing like a PhD thesis. This did add further complexity to the detection of the use of AAG tools.

A: Can you talk to me please about your PhD student who you believe used AAG tools? How did you discover this?

S: Sorry to say, it took me more than 2 years until we solved this mystery! The student had developed the first research proposal in 6 months and that proposal was rejected. Then, the student was given another six months extension. With big support from the supervisory team, the proposal managed to pass after 14 months in total. This obviously raised initial concerns about the student’s ability.

A: During these months, did you not notice the use of AAG tools?

S: No, but we observed an inconsistent writing style, and we thought that the student had some sort of logical thinking problems. We were not aware of AAG tools.

A: Can you please describe what you mean by inconsistent writing style?

S: Okay, each paragraph had several phrases. Some of these phrases were really good and related to the topic and others were not. In general, these phrases did not present a clear idea. We thought that the student was collecting information from different papers without enough understanding. I discussed the problem of the student’s writing with other staff, but none suspected use of AAG tools.

A: What did they think was happening?

S: Some blamed me for accepting the student for supervision, others questioned how such a proposal passed via the panel? Another claimed that the quality of students is decreasing. Anyway, we (the supervisors) put together a plan for supporting the student to develop the first research chapter.

A: What was that plan?

S: The plan was providing close assistance for the student, explaining more about the research topic, extra communication and meetings. In return, we figured clearly the student should progress.

A: Then what happened?

S: As part of this close assistance, I reviewed the student's draft writing, phrase by phrase, with the student. So, I totally figured out the student's abilities. The student did not understand most of their own writing and could not explain the writing they were presenting.

A: Again, did you suspect that the student was using AAG tools?

S: Absolutely not, because we did not know about these tools or even imagine they existed for writing a thesis. I forgot to mention that, during the meetings, the student was trying to direct me to secondary issues or unrelated topics. Also, the student was trying to consume the time of the meeting by asking irrelevant questions. The student was very careful to not let us discuss the research. In other words, I realised later that they were trying to distract me away from the written work. I was noting this, but I was not getting an explanation.

A: Then?

S: Then, we asked the student to complete the first research chapter (planned to be the first journal paper submission), without the previous close assistance, which should have kept them busy for several weeks. However, in only two months the student claimed to have completed not only the first chapter but also said they had finalised another two research chapters. Then, two weeks later the student submitted another chapter for review.

A: You mean that, the student prepared three chapters within two months?

S: Yes, the student provided three new chapters plus finalising the first chapter! I reviewed the first chapter and informed the student that the work was not at the level needed for a PhD. However, the student claimed that they had different research opinions to the supervisors and that we did not understand their research views. To break the deadlock, we (the supervising team) decided to give the student another chance to revise the first chapter based on our feedback and then submit it to a related recognised journal to get good external feedback, since they would not accept the supervisors' feedback. We did not bother to review the extra chapters at that time.

A: Then?

S: After a few months and significant further reviews from the supervisory team, which the student only partially followed, the first research chapter was submitted to a Q1 journal. While that was under review we started to look at the student's other chapters (the three chapters that were quickly developed previously). At the first look through one of them, I got a feeling that this writing is not human writing! This was a literature review chapter.

A: Could you please tell me what you noticed at this time, but not before?

S: I identified the fake writing this time because the student's writing did not contain any input from the supervisors. All sections, concepts, paragraphs, headings and sub-headings were written by the student. I believe the reason why we could not detect this issue earlier was because the student's writing was edited and changed several times by us through the oral and written feedback, which was continuously given on the first chapter as we tried to get the student started on the right path for their thesis. Additionally, the student was presenting their writing in pieces (section by section, in other words few paragraphs by few paragraphs) and got the related feedback, then rewrote them and got more feedback and so on. The feedback was given for each section separately.

A: This is a good point. Editing of AAG writing converts it to a distortion of the cyborg's writing and makes it difficult to recognize who produced it, a robot or human.

S: Yes, particularly, since it was the supervisors who edited it.

A: What was the result from the journal article review?

S: The reviewers requested major revisions due to the same technical issues that the supervisors had raised. But they did not raise any AAG issues or significant writing issues. Of course, the text had been extensively edited several times so this masked the likely original generation from AAG. However, because by this stage we had become increasingly suspicious that the student was using AAG tools to generate their work, we of course contacted the editors, explained our concerns and withdrew the paper from further consideration, with apologies.

A: What did you observe in the next chapters that made you think about the possibility of AI tools although you were not aware about them before?

S: The first figure in the next chapter was confusing. It consisted of three parts a, b and c. Each one of these parts came from different articles. The description of each part was very clear and well written but there was no relationship between those three parts; there was no theory nor concept that could connect them, except the similarity of the keywords. I thought that whoever connected those three parts together had minimal research sense or had not read or understood the cited research papers properly. On the other hand, the descriptions of each part were professionally written. After reviewing the whole manuscript, I found that the student always collected some phrases about one idea from different resources then added them together to make one paragraph. These paragraphs often did not make any sense although they had correct information, and they were often irrelevant to the topic. Also, there were always abrupt shifts between the ideas without logical transitions.

A: This description is clear for me. Did the other supervisors notice this issue?

S: Once I raised these concerns the rest of the team investigated my suspicions. The first language English member of the team indicated that the speed with which the student produced these chapters and the sentences within them added some weight to my suspicions. They commented on one paragraph as an example where the paragraph didn't make sense and also couldn't possibly be the student's own writing because the English words and phrases were very complex, but then the phrases seemed to just be thrown together in odd ways. Another supervisor stated that the number of references used was too many and that it would have been impossible for the student to read them all in the time in which they produced the chapter. Another supervisor noted that some of the references used were very old and obscure, and that some of the references were not in English or in the student's first language so could not have been read by the student.

A: This description agrees well with the article generator products. However, is it possible that the student is innocent and did not use the AAG tools?

S: Well, in the end we did not use these suspicions as the reason for the candidate's eventual expulsion as there were other concerns as well, so we cannot say that we conclusively proved that the student used AAG tools. But the primary evidence was that although the concerns about the writing were extensively discussed with the student (without specifically accusing them of using AAG), they continued to be evasive about the questions and could not give any reasonable explanations for the rapid production of the writing, and the issues around coherence, relevance and references.