



Jack Watson: Addressing Contract Cheating at Scale in Online Computer Science Education

Rocko Graziano
Georgia Institute of
Technology
Atlanta, GA, USA
rockograziano@gatech.edu

Qiuyue Xue
Georgia Institute of
Technology
Atlanta, GA, USA
qiuyue@gatech.edu

Diego Vacanti
Georgia Institute of
Technology
Atlanta, GA, USA
dvacanti3@gatech.edu

Deepak Srikanth
Georgia Institute of
Technology
Atlanta, GA, USA
deepaksrikanth@gatech.edu

David Benton
Georgia Institute of
Technology
Atlanta, GA, USA
dbenton@gatech.edu

P. Tim Miller
Georgia Institute of
Technology
Atlanta, GA, USA
ptim.miller@gatech.edu

Pepper Miller
Georgia Institute of
Technology
Atlanta, GA, USA
pepper@gatech.edu

Thad Starner
Georgia Institute of
Technology
Atlanta, GA, USA
thad@gatech.edu

Sarthak Wahal
Georgia Institute of
Technology
Atlanta, GA, USA
sarthak.wahal@gatech.edu

Nick Larsen
Georgia Institute of
Technology
Atlanta, GA, USA
nicklarsen@gatech.edu

Khushhall Chandra Mahajan
Georgia Institute of
Technology
Atlanta, GA, USA
khushhall@gatech.edu

ABSTRACT

Cheating has always been a problem for academic institutions, but the internet has increased access to a form of academic dishonesty known as contract cheating, or “homework for hire.” When students purchase work online and submit it as their own, it cannot be detected by commonly-used plagiarism detection tools, and this troubling form of cheating seems to be increasing.

We present an approach to addressing contract cheating: an AI agent that poses as a contractor to identify students attempting to purchase homework solutions. Our agent, Jack Watson, monitors auction sites, identifies posted homework assignments, and provides students with watermarked solutions that can be automatically identified upon submission of the assignment.

Our work is ongoing, but we have proved the model, identifying nine cases of contract cheating through our techniques. We are continuing to improve Jack Watson and further auto-

mate the monitoring and identification of contract cheating on online marketplaces.

Author Keywords

plagiarism, contract cheating, academic integrity

ACM Classification Keywords

K.3.1 Computer Uses in Education: Distance learning

INTRODUCTION

As long as there have been schools, there has been cheating. It is a dark truth of human nature that we will game the system for individual gain. Cheating is a serious threat to universities, whose reputation depends on producing well-qualified future professionals; to future employers, who depend on employees who have the level of knowledge they claim and do not shirk their responsibilities; and consumers of the products these students may produce. When there is a lapse in academic integrity, students are no longer being evaluated on a level playing field. Honest students may perceive that they are struggling where others seem to cope easily with assignments. This situation may be demotivating to students who are most engaged in the work. If unaddressed, these effects will rightly erode students’ trust in their educational institutions. If cheating is allowed to become widespread, the university’s reputation may be damaged, and the value of its education will ultimately decrease.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

L@S '19 June 24–25, 2019, Chicago, IL, USA

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6804-9/19/06.

DOI: 10.1145/3330430.3333665

On average, workers with a Masters degree in Computer Science earn \$18k per year more than those with a Bachelor's degree in CS as calculated using the statistics at Payscale.com [11]. Students are therefore highly motivated to obtain a degree and may turn to cheating if they feel they cannot earn it honestly.

Contract cheating, i.e., when a student purchases original work to be submitted as their own, is especially troubling for schools, because it cannot be identified by common plagiarism detection tools. The internet and the gig economy have created a "homework for hire" marketplace for contract cheating. Experts on any subject are available, and contract cheating appears to be on the rise[9]. Schools need a way to address this threat to academic integrity.

Given the challenges of identifying contract cheating, we propose an automated approach: an AI agent that poses as a contractor on auction sites to identify students who are trying to cheat.

In our ongoing work, we aim to address the following research questions:

1. Can monitoring of contract cheating marketplaces be automated?
2. Considering the breadth of projects posted for hire online, many of them legitimate commercial postings, is it possible to identify which are contract cheating?
3. When a project has been identified as homework, is it possible to automate the collection of additional information to identify the contract is for a specific assignment?
4. Can an AI agent win auction contracts from students attempting to purchase homework solutions?

BACKGROUND

Generally, the literature considers only paid work to be contract cheating, but there are some studies that include unpaid work by friends and family [8, 3]. We define contract cheating as paid work completed by a third party for submission by a student. As it involves submitting unattributed work, contract cheating is a form of plagiarism. Contract cheating is an especially troubling form of cheating, because it is deceptive, premeditated, and difficult to detect. Surveys of student attitudes of cheating show that students are aware that contract cheating is wrong, and many students consider it more serious than other forms of cheating [6, 14].

Prevalence of Contract Cheating

Despite students' views that contract cheating is a more serious offense, recent self-report surveys suggest it is increasing in prevalence. Most studies on the rate of contract cheating consist of self-report surveys, but there are some obvious limitations to self-reporting of illicit behavior. Reassuringly for the use of self-report surveys, Gardner et al. [7] showed that admissions of cheating correlated positively with actual cheating, as long as there was no penalty for the admission.

A review of the literature [9] found a historic average of 3.52% of students report contract cheating, and found that fitting a

simple linear regression model showed a significant positive relationship over time. Surveys since 2015 had an average 15.7% of students self-report contract cheating. While the methods used to study the prevalence of contract cheating are imperfect, it is likely that contract cheating is on the rise. Even if current rates were 3.52%, it would represent 7 million students worldwide, and true rates may be significantly higher based on more recent surveys.

Contract Cheating and Online CS Education

Studies have shown a disproportionate number of academic integrity violations by computer science students. Over a ten year period, CS students at Stanford University accounted for 37% of academic integrity violations, even though they only made up 6.5% of the student body [13]. In surveys performed by Alam [2], students self-reported significantly more plagiarism in programming assignments than essays and other types of assignments. Computer science education has also been linked to contract cheating since the term was coined by Clarke and Lancaster [4], in their 2006 study of student postings on the programming auction site RentACoder.

The scale of online education, with hundreds of students per class, and the lack of in-person interaction leaves teaching staff with less personal knowledge of individual students' abilities. This lack of in-person interaction may make it harder for staff to assess whether an assignment was completed by the student. In online CS education, autograders may be used for some types of assignments, removing human graders from the process and precluding grader identification of contract cheating.

PRIOR WORK

Clarke and Lancaster [4] found that 99 of 803 (12.3%) projects posted to the auction site RentACoder during a three week period in 2005 were highly likely to be contract cheating. In most cases, they were able to identify the originating institution of the assignment. In a follow-up to their previous study, Clark and Lancaster [5] evaluated a corpus of 14,438 identified attempts to cheat, collected between 2005 and 2012. That number does not represent all attempts at contract cheating during the period, even from the sites that were included in the study, as the methodology was manual identification of a subset of posts on those sites. This corpus was used to estimate the size of the contract cheating market on a single site, freelancer.com, which averaged £718,478 of contracts annually. The authors noted difficulty in attributing assignment specifications to an institution or instructor, and suggested changes that teaching staff can use to make assignments more searchable and identifiable.

O'Malley and Roberts [10] proposed a method they called a "cheat trap solution," and shared a case study of the successful use of the method. The authors were alerted to a contract cheating post for one of their assignments on an auction site. Subsequently, they bid on the project, won the contract, and provided the student with a cheat trap solution, which included easily-identifiable features. Graders were alerted to look for these features, and the student was identified upon submission of the cheat trap solution. We also use identifiable features, a

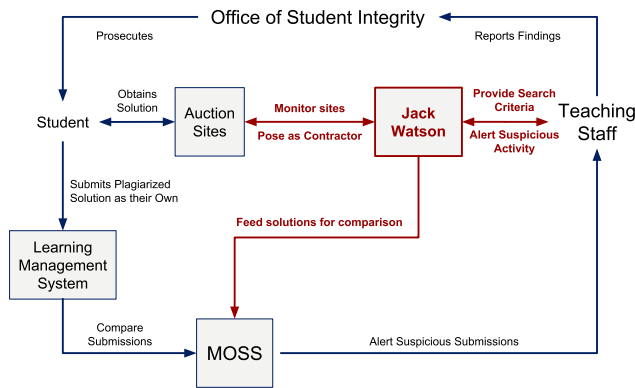


Figure 1. The red is the scope of the Jack Watson agent, which poses as a contractor on auction sites to catch students engaged in contract cheating.

technique which we call watermarking, to identify students engaged in contract cheating.

JACK WATSON

To address contract cheating in online education, we introduce Jack Watson, an AI agent designed to combat contract cheating. Jack Watson monitors the internet for contract cheating projects, communicates with buyers, and provides watermarked solutions. It also includes tools to help teaching staff monitor specific keywords and repeat buyers.

To understand how Jack Watson interacts with commercial auction sites, it is helpful to understand a typical workflow on these sites. First, a buyer posts information about their project, and contractors bid on the work. The buyer can then communicate with bidding contractors to clarify the details of the project, and award the project to the contractor of their choice. On some sites, parts of the project posting and bidding process are publicly visible. This openness creates an opportunity for monitoring projects for identification of possible contract cheating.

Jack Watson monitors this publicly-available information continuously and saves project details and buyer information for posted projects. Projects are classified based on the likelihood that the project is attempted contract cheating. Suspect projects are flagged for human review.

Teaching staff have access to a dashboard where they can enter assignment details to monitor for specific assignments. Assignments are identified by a date range and a set of keywords. Assignment instructions may be uploaded for automatic keyword extraction, or keywords may be selected manually. Jack Watson automatically bids on suspect projects that fall within the given date range and contain the selected keywords.

Once a project has been identified as a likely homework assignment, and a bid has been placed, Jack Watson chats directly with the student. There are two goals of this interaction: first, to positively identify whether the project is the assignment in question, and second, to convince the student that Jack Watson is sufficiently experienced to provide a solution. Students engaged in contract cheating may post only minimal information

I am looking for some help on bi-directional a* search.

I am familiar with that topic. Could you please direct me to your problem statement.

Do you want me to cut and paste it here?

feel free to contact me here: jackwatson@email.com

Ok I am sending a pdf with the details to your email

yes, please pass it on

Figure 2. Example Student communication (left); Jack Watson response (right) is generated by DialogFlow and intended to solicit information necessary to identify project as homework assignment.

for a project, being wary of posting specifics publicly. Project bid negotiations are private, and students are more open about the details in this setting. Our agent encourages the student to share information that can be used to positively identify an assignment (see Figure 2).

When Jack Watson bids on a project, relevant teaching staff are notified of the project. Staff may monitor the progress of negotiations through the Jack Watson interface and may take over communication from the AI agent at any point.

If Jack Watson is awarded the project, teaching staff is prompted to provide a working solution that has been watermarked to make it identifiable upon submission, without raising suspicion of the student, similar to the technique described by O'Malley and Roberts [10]. While some information about the buyer is available to the winning bidder, it is typically not enough to identify them. Generally, the identity of the student will remain unknown until the watermarked solution is submitted.

Identification of watermarked solutions is easily incorporated into the existing assignment grading workflow. It is common practice to check programming assignment submissions for plagiarism using automated tools. Specialized tools, such as MOSS [1] or JPlag [12], are used to compare all submitted assignments and identify similarities between submissions. Any watermarked solutions provided by Jack Watson are submitted to the plagiarism detection software along with student submissions. If there are significant matches with the watermarks, it is strong evidence of contract cheating. Any student identified in this way is referred for investigation and discipline.

RESULTS TO DATE

Our experience has shown that it is indeed possible to automate continuous monitoring of auction sites. Jack Watson has been monitoring sites for more than a year and has collected public information on millions of projects. Some projects are easily identified as contract cheating requests posted by students. In rare cases, it may be possible to positively identify a specific

assignment from only the public project posting. Many other projects are clearly commercial or non-educational in nature.

We have also validated the use of watermarked solutions to identify students. Nine cases of contract cheating have been identified so far and referred for discipline after students have submitted watermarked solutions.

One hope for Jack Watson is that it will act as a disincentive, lowering the rates of attempted contract cheating. For several semesters, we have publicized to students some basic information about our research. Over this time, there has been a decrease in the number of easily identifiable contract cheating projects, especially in the class for which Jack was first developed. While there are a number of possible factors, including new efforts to raise student awareness regarding plagiarism, we believe Jack Watson may already be acting as a disincentive for students to cheat.

FUTURE WORK

While initial results are promising, we are engaged in ongoing work to address our research questions and improve the effectiveness of Jack Watson. We have proved the concept with human involvement. While human oversight will always be required, we are working to decrease the agent's reliance on human intervention. It is an open challenge to develop a persona that students will trust enough to hire.

We plan to monitor additional auction sites. The design of certain sites, such as fiverr.com, reverses the bidding process: contractors post services that they offer, and buyers contact them. This removes the public exposure of student-posted project specifications, which precludes public monitoring. It may be possible to create honey pot contractor accounts on these sites.

Currently, Jack Watson is focused on computer science education, but the technology may have applications in other subjects as well. We expect to add more classes to monitor for specific assignments and eventually hope to add other schools.

CONCLUSION

Contract cheating is a serious problem for universities, and existing plagiarism detection tools are not effective in identifying it. To address this issue, we have created an AI agent, Jack Watson, to identify students who are actively engaged in contract cheating. We aim to decrease the prevalence of contract cheating directly by catching students, and indirectly by creating a disincentive for students through awareness of Jack Watson's monitoring. Our results are promising, with a number of cases successfully identified and an apparent decrease in cheating attempts. We are eager to continue to improve Jack Watson, and to decrease rates of contract cheating.

ACKNOWLEDGEMENT

This work was conducted over a two-year period in Georgia Tech's OMSCS program. The authors would like to recognize the following for their contributions to Jack Watson: College of Computing: Tanmay Binaykiya; OMSCS: Brent Dornier, Sachin Garg, Chris Hedenberg, Daniel Kirel, Harsh Kohli, Jesus Magana, Matt Martz, Michal Mati, Girish Mohandass, Michael O'Connor, Brandon Smith, Chris Smith, Nathan Susanj, Alicia Tsai, and Mark Wilson.

REFERENCES

1. Alex Aiken. 1994. Moss: A System for Detecting Software Similarity. <http://theory.stanford.edu/~aiken/moss/>. (1994). Accessed: 2019-04-01.
2. Lubna S Alam. 2004. Is plagiarism more prevalent in some forms of assessment than others. In *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference*. 48–57.
3. Tracey Bretag, Rowena Harper, Michael Burton, Cath Ellis, Philip Newton, Pearl Rozenberg, Sonia Saddiqui, and Karen van Haeringen. 2018. Contract cheating: A survey of Australian university students. *Studies in Higher Education* (2018), 1–20.
4. Robert Clarke and Thomas Lancaster. 2006. Eliminating the successor to plagiarism? Identifying the usage of contract cheating sites. In *proceedings of 2nd international plagiarism conference*.
5. Robert Clarke and Thomas Lancaster. 2013. Commercial aspects of contract cheating. In *Proceedings of the 18th ACM conference on Innovation and technology in computer science education*. ACM, 219–224.
6. Guy J Curtis and Lucia Vardanega. 2016. Is plagiarism changing over time? A 10-year time-lag study with three points of measurement. *Higher Education Research & Development* 35, 6 (2016), 1167–1179.
7. William M Gardner, James T Roper, Claudia C Gonzalez, and Royce G Simpson. 1988. Analysis of cheating on academic assignments. *The Psychological Record* 38, 4 (1988), 543–555.
8. Zaigham Mahmood. 2009. Contract cheating: a new phenomenon in cyber-plagiarism. *Communications of the IBIMA* 10, 12 (2009), 93–97.
9. Phil Newton. 2018. How common is commercial Contract Cheating in Higher Education?. In *Frontiers in Education*, Vol. 3. Frontiers, 67.
10. Michael O'Malley and Tim Sean Roberts. 2012. Plagiarism on the rise? Combating contract cheating in science courses. *International Journal of Innovation in Science and Mathematics Education* 20, 4 (2012).
11. PayScale. 2019. PayScale. <https://www.payscale.com/>. (2019). Accessed: 2019-04-01.
12. Lutz Prechelt, Guido Malpohl, and Michael Philippsen. 2002. Finding plagiarisms among a set of programs with JPlag. *J. UCS* 8, 11 (2002), 1016.
13. Eric Roberts. 2002. Strategies for promoting academic integrity in CS courses. In *32nd Annual Frontiers in Education*, Vol. 2. IEEE, F3G–F3G.
14. Liora Pedhazur Schmelkin, Kim Gilbert, Karin J Spencer, Holly S Pincus, and Rebecca Silva. 2008. A multidimensional scaling of college students' perceptions of academic dishonesty. *The Journal of Higher Education* 79, 5 (2008), 587–607.