Detecting and avoiding plagiarism in the era of technological developments and AI

Mohammad Khalil

Plagiarism in Higher Education - awareness, threats and consequences

26-27 Feb 2024

Podgorica, Montenegro

mohammad.khalil@uib.no

@TUMohdKhalil
The images used in the presentations belong to their respective owners, and I acknowledge the original copyright and licensing attributed to them.
menti.com

Use code: 3661 9376
Rules

1. No cheating!!!
2. No cell phones!
3. No gum!

Cheating <> Proctoring
Chinese schoolgirl caught using robot to write her homework—now everybody wants one

South China Morning Post | Published: 9:14am, 19 Feb, 2019
Two research articles for insights
Will ChatGPT Get You Caught? Rethinking of Plagiarism Detection

Mohammad Khalil & Erkan Er

Conference paper | First Online: 09 June 2023

837 Accesses | 19 Citations

Part of the Lecture Notes in Computer Science book series (LNCS, volume 14040)
La matematica dell’originalità

L’unico metodo efficace per scoprire se un testo è stato elaborato da ChatGPT è chiederglielo a ChatGPT.

di Mohammad Khalil e Erkan Er

In un recente articolo intitolato The College Essay Is Dead (Il saggio universitario è morto, N.D.T.) pubblicato su The Atlantic, Stephen Marche esprime preoccupazione riguardo all’utilizzo di ChatGPT nella creazione di testi di alta qualità, spesso di natura accademica, sfruttando le capacità di elaborazione del linguaggio naturale dei chatbot.

Al contrario, Stockel-Walker sottolinea come ChatGPT possa essere di aiuto agli studenti in diversi modi, per esempio nello scrivere saggi, risolvere compiti, creare script e nella valutazione delle loro prestazioni.

I sistemi anti-plagio — Il plagio è il presentare il lavoro o le idee di qualcun altro come propri, senza indicarne l’autore o la fonte. Un fenomeno che non riguarda solamente il testo, ma può estendersi anche alle immagini.

Oggi il dilagante problema del plagio online, soprattutto nei compiti e negli elaborati scritti, costituisce una delle principali sfide che la scuola deve affrontare.

I più diffusi sistemi anti-plagio, secondo quanto riportato da alcune fonti, sono Turnitin e iThenticate, due software sviluppati dalla stessa azienda, iParadigms LLC. Dal 1997 sono stati sempre più utilizzati nelle istituzioni scolastiche, in particolare dalle università.

Il loro funzionamento è piuttosto semplice: con un click sono in grado di dire se un testo o parte di esso è stato copiato da internet o da un altro testo disponibile in rete. Confrontano il compito dello studente con i testi contenuti nel database (materiali disponibili sul web, testi accademici, elaborati, tesi, articoli scientifici, papers) e producono un report dettagliato indicando la percentuale di originalità. Ma quanto sono affidabili questi sistemi?

Metodologia della ricerca — In un nostro studio abbiamo presentato i risultati di un’analisi di plagio effettuata su alcuni contenuti generati dall’intelligenza artificiale. In particolare abbiamo condotto un’analisi quantitativa, in cui il materiale generato da un chatbot è stato misurato attraverso gli indici di originalità prodotti da due software anti-plagio.

Per ottenere un campione rappresentativo abbiamo proposto 50 diversi argomenti e incaricato ChatGPT di redigere un saggio di 500 parole su ognuno di essi. Ciascun output è stato poi convertito in un testo e salvato in un file separato, come se fosse stato redatto da uno studente.

La prima metà di questo ma-
Study Objective

• Investigate LLMs- OpenAI ChatGPT- for academic honesty and plagiarism check from an academic perspective
  
  • Evaluate the originality of 50 essays generated by ChatGPT on various topics.
The Mechanism Behind ChatGPT
Software used for plagiarism check

logos copyrights of turnitin
The 50 essays

- We asked chatgpt 3 to generate 50 essays on different topics
- Each was saved separately
- 50% were sent to Turnitin
- The other 50% were sent to iThenticate
## Results (similarity check)

<table>
<thead>
<tr>
<th>Essay Title</th>
<th>Similarity</th>
<th>Date and Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>20- cloud storage.docx</td>
<td>7%</td>
<td>Jan 23, 2023 1:59:40 PM</td>
</tr>
<tr>
<td>6 - MOOCs.docx</td>
<td>3%</td>
<td>Jan 23, 2023 1:59:39 PM</td>
</tr>
<tr>
<td>13- SNA.docx</td>
<td>14%</td>
<td>Jan 23, 2023 1:59:38 PM</td>
</tr>
<tr>
<td>21- learning theories.docx</td>
<td>14%</td>
<td>Jan 23, 2023 1:59:38 PM</td>
</tr>
<tr>
<td>19- cloud computing.docx</td>
<td>19%</td>
<td>Jan 23, 2023 1:59:38 PM</td>
</tr>
<tr>
<td>10- machine learning.docx</td>
<td>30%</td>
<td>Jan 23, 2023 1:59:37 PM</td>
</tr>
<tr>
<td>24- constructivism.docx</td>
<td>2%</td>
<td>Jan 23, 2023 1:59:37 PM</td>
</tr>
</tbody>
</table>

(n=7)
Robots are machines that are capable of performing tasks autonomously or semi-autonomously. They have been used for a variety of purposes throughout history, from manufacturing and assembly to search and rescue operations. Today, robots are becoming increasingly advanced and are being used in a wide range of industries, from healthcare and transportation to entertainment and education.

One of the most common uses for robots is in manufacturing and assembly. In these industries, robots are used to perform tasks that are too dangerous or difficult for humans to do. For example, robots can be used to weld, paint, and assemble products at high speeds and with a high degree of precision. They can also be used to handle hazardous materials and operate in environments that are too hot, cold, or dirty for humans to work in.

Another area where robots are increasingly being used is in healthcare. Medical robots can be used to assist surgeons in performing complex procedures, such as brain surgery or heart surgery. They can also be used to help patients with physical therapy and rehabilitation. In addition, robots are being used to perform routine tasks in hospitals, such as delivering medication and cleaning patient rooms.

Robots are also being used in transportation. Self-driving cars and drones are examples of robots that are being used to make transportation safer and more efficient. Self-driving cars can reduce the number of accidents caused by human error, and drones can be used to deliver packages and perform search and rescue operations.

In addition to these practical uses, robots are also being used in entertainment and education. For example, robots are being used to create interactive exhibits in museums and theme parks, and to help teach children about science, technology, engineering, and math (STEM).

Despite all the benefits that robots bring, there are also concerns about their impact on society. For example, there are concerns that robots will take jobs away from humans and lead to increased unemployment. There are also concerns about the safety of robots, as they may malfunction and cause harm to humans.

However, it is important to note that robots are not designed to replace human workers, but to work alongside them and make their jobs easier and safer. Moreover, with the advance in technology, many jobs that were considered to be impossible to automate are being done by robots, and this has increased productivity and efficiency.

In conclusion, robots are machines that are capable of performing a wide range of tasks, from manufacturing and assembly to healthcare and transportation. While there are concerns about their impact on society, robots have the potential to make our lives easier and safer. As technology continues to advance, we can expect to see even more uses for robots in the future.

An example of 5% similarity, Topic of ‘Robots’

Plagiarism Software: Turnitin
Learning theories are frameworks that explain how individuals acquire and process information. They provide a way for researchers and educators to understand how people learn, and to design educational programs and interventions that are effective. There are several different learning theories, each of which provides a unique perspective on the learning process.

One of the most widely-known learning theories is classical conditioning, proposed by Ivan Pavlov. Classical conditioning is a type of learning that occurs when a neutral stimulus is paired with a stimulus that naturally evokes a response. For example, the sound of a bell (neutral stimulus) can be paired with food (natural stimulus) to make a dog salivate (response) when it hears the bell. This theory suggests that learning occurs through the repeated pairing of stimuli.

Another influential learning theory is operant conditioning, proposed by B.F. Skinner. This theory states that learning occurs when behavior is reinforced or punished. Reinforcement is something that increases the likelihood that a behavior will be repeated, while punishment is something that decreases the likelihood of a behavior being repeated. For example, if a student receives praise for turning in a homework assignment on time, they are more likely to repeat that behavior in the future.

Cognitive learning theory is a perspective that emphasizes the role of mental processes in learning. This theory, proposed by Jean Piaget, suggests that learning is an active process where individuals construct their own understanding of the world by organizing and reorganizing information. This theory emphasizes the importance of understanding and addressing the cognitive processes that are involved in learning.

Social learning theory, proposed by Albert Bandura, suggests that individuals learn by observing and imitating the behavior of others. This theory focuses on the role of observational learning, which occurs when people observe and imitate the actions of others. This theory emphasizes the importance of role models, and the influence of the social context on learning.

Finally, constructivism is a learning theory that emphasizes the role of the learner in actively constructing their own understanding of the world. Constructivism suggests that people learn best when they are actively engaged in the learning process, and when they are able to connect new information to what they already know. This theory emphasizes the importance of hands-on experiences, exploration, and discovery in learning.

In conclusion, learning theories are frameworks that explain how individuals acquire and process information. There are several different learning theories, each of which provides a unique perspective on the learning process. These theories include classical conditioning, operant conditioning, cognitive learning theory, social learning theory, and constructivism. Understanding these theories can help educators and researchers to design effective educational programs and interventions that meet the needs of learners.
An example of 64% similarity, Topic of ‘Laws of physics’

Plagiarism
Software: Turnitin
## Table 1. iThenticate® Plagiarism check results (n = 25 essays)

<table>
<thead>
<tr>
<th>Essay topics</th>
<th>Essay count (%)</th>
<th>Similarity score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud storage; Massive open online courses (MOOCs); constructivism; Robots; use of smartphones; Internet revolution; unsupervised machine learning; creativity; assessment in education; Natural Language Processing (NLP); Driving schools; use of chatbots in education; Technology-Enhanced Learning; self-regulated learning; online banking; leadership; spam emails; hybrid learning</td>
<td>17 (68%)</td>
<td>&lt; 10%</td>
</tr>
<tr>
<td>Social Network Analysis; learning theories; cloud computing; classification in machine learning; marketing plans</td>
<td>5 (20%)</td>
<td>10–20%</td>
</tr>
<tr>
<td>Machine learning; prediction; clustering</td>
<td>3 (12%)</td>
<td>20–40%</td>
</tr>
<tr>
<td>None</td>
<td>0 (0%)</td>
<td>&gt; 40%</td>
</tr>
<tr>
<td><strong>Total and Average</strong></td>
<td>Total (n = 25)</td>
<td>Average (8.76%)</td>
</tr>
</tbody>
</table>
Plagiarism check using Turnitin

<table>
<thead>
<tr>
<th>Essay topics</th>
<th>Essay count (%)</th>
<th>Similarity score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergartens; Cultures of the Middle East and South America; Hybrid and blended teaching; Educational measurement; Difference of jobs in California and New York; Flipped vs traditional lecturing; Clustering and association rule mining; Psychologists and psychiatrists; Differential equations; PhD (Doctoral holder); Good teacher; Respiratory systems</td>
<td>12 (48%)</td>
<td>&lt; 10%</td>
</tr>
<tr>
<td>Clustering algorithm; C# and Java; Data science and machine learning; Object Oriented Programming; Computer science and computer engineering; Organic chemistry</td>
<td>6 (24%)</td>
<td>10–20%</td>
</tr>
<tr>
<td>Child usage of screens; Learning Analytics and Educational Data Mining; Deep learning; Logistic regression; Global warming; Data structure</td>
<td>6 (24%)</td>
<td>20–40%</td>
</tr>
<tr>
<td>Laws of physics</td>
<td>1 (4%)</td>
<td>&gt; 40%</td>
</tr>
<tr>
<td><strong>Total and Average</strong></td>
<td><strong>Total (n = 25)</strong></td>
<td><strong>Average (13.72%)</strong></td>
</tr>
</tbody>
</table>
Reverse Engineering

92% is true positive cases
When A.I. Chatbots Hallucinate

By Karen Weise and Cade Metz
Karen Weise reported this story from Seattle and Cade Metz reported from San Francisco.

Published May 1, 2023  Updated May 9, 2023

When did The New York Times first report on “artificial intelligence”?

According to ChatGPT, it was July 10, 1956, in an article titled “Machines Will Be Capable of Learning, Solving Problems, Scientists Predict” about a seminal conference at Dartmouth College. The chatbot added:
Yes, this text was generated by a chatbot (me, ChatGPT).
Interesting Remarks, Paper 1
Remarks

- Students may possibly use ChatGPT to complete essay-type assignments without getting caught.
- Topics matter: factual vs interpretative.
- We show evidence that plagiarism with ChatGPT is already a pressing concern requiring attention.
- Popular anti-plagiarism software companies are in challenge.
Testing of detection tools for AI-generated text

Debora Weber-Wulff¹, Alla Anohina-Naumeca², Sonja Bjelobaba³, Tomáš Foltýnek⁴, Jean Guerrero-Dib⁵, Olumide Popoola⁶, Petr Šigut⁴ and Lorna Waddington⁷
Interesting Remarks, Paper 2
Remarks

- “Detection tools for AI-generated text do fail, they are neither accurate nor reliable (all scored below 80% of accuracy and only 5 over 70%)” (pp.25)

- AI Plagiairism detection tools diagnose human-written documents as AI-generated (false positives) context

- “easy solution” for detection of AI-generated text does not (and maybe even could not) exist! - Weber-Wulff et al. 2023.
Questioning the Reliability of AI detection tools
Guidance on AI Detection and Why We’re Disabling Turnitin’s AI Detector

Posted by Michael Coley on Wednesday, August 16, 2023 in Announcements, News.

In April of this year, Turnitin released an update to their product that reviewed submitted papers and presented their determination of how much of a paper was written by AI. As we outlined at that time, many people had important concerns and questions about this new tool, namely how the product exactly works and how reliable the results would be. After several months of using and testing this tool, meeting with Turnitin and other AI leaders, and talking to other universities who also have access, Vanderbilt has decided to disable Turnitin’s AI detection tool for the foreseeable future. This decision was not made lightly and was made in pursuit of the best interests of our students and faculty.

When Turnitin launched its AI-detection tool, there were many concerns that we had. This feature was enabled for Turnitin customers with less than 24-hour advance notice, no option at the time to disable the feature, and, most importantly, no insight into how it works. At the time of launch, Turnitin claimed that its detection tool had a 1% false positive rate (Chechitelli, 2023). To put that into context, Vanderbilt submitted 75,000 papers to Turnitin in 2022. If this AI detection tool was available then, around 750 student papers could have been incorrectly labeled as having some of it written by AI. Instances of false accusations of AI usage being leveled against students at other universities have been widely reported over the past few months, including multiple instances that involved Turnitin (Fowler, 2023; Klee, 2023). In addition to the false positive issue, AI detectors have been found to be more likely to label text written by non-native English speakers as AI-written (Myers, 2023).
Detecting AI may be impossible. That’s a big problem for teachers.

Turnitin has acknowledged a reliability problem with AI cheating-detection software used on 38 million student papers. Computer scientists warn we may never be able to reliably detect AI.
Bypass Turnitin AI detection

Start free trial

If there is a detector, we bypassed it

Turnitin  GPTZero  Originality AI  Copyleaks  Writer  ChatGPT Content Detector  GPTRadar  Winston AI
Content At Scale  Corrector  ZeroGPT  Crossplag  Sapling
How plagiarism can be addressed with LLMs?

- Teacher/tutors/instructors are advised to:
  - Give assignments that go beyond the basics and foster active engagement and critical thinking,
  - Inform students of the limitations of ChatGPT and the potential consequences of relying merely on it,
  - Underline the importance of academic integrity and ethical behaviour and provide clear guidelines and expectations for students in syllabus
How plagiarism can be addressed with LLMs?

• Students/pupils/learners are advised to

  • Take advantage of this technology as a means to improve their competencies and learning, but not as a substitute for original thinking and writing,

  • Be aware of the proper and ethical use of ChatGPT in their courses and the consequences of solely relying on it for academic integrity.
How plagiarism can be addressed with LLMs?

- Institutions are advised to
  - Get familiarised with the potentials of large language models in education and open communication channels to discuss transparently with involved stakeholders, including researchers and IT support,
  - Create and implement clear policies and guidelines for the use of AI tools, such as ChatGPT,
  - Offer training and resources for students, faculty, and staff on academic integrity and the responsible use of AI tools in education.
As teachers, What else can we do about it?
More actions

• Suspicious?

• Check writing style

• Compare to other work by the same student (e.g., style, tone, level of writing)

• Look for inaccuracies in sources, arguments, facts

• Look for hallucinations and strange results (LLMs are not search engines!)
Detecting and avoiding plagiarism in the era of technological developments: The case of AI

Thank you!

Plagiarism in Higher Education - awareness, threats and consequences

26-27 Feb 2024

Podgorica, Montenegro

mohammad.khalil@uib.no

@TuMohdKhalil