International Journal of
Science, Architecture, Technology and Environment

Advances in AI-Generated Text Detection: A Systematic Review

Dr. S. V. Viraktamath[1], Mahmad Sohel[2]*, Bhogeshrao[3], Anita Mallangoudra[4]

[1]Department of Electronics and Communication Engineering, S D M College of Engineering and Technology Dharwad, Karnataka State, India

**Corresponding author, sohelmudgal85@gmail.com*

DOI: https://doi.org/10.63680/ijsate0126061.054

Abstract

Advanced language models can now create writing that looks very similar to human text, making it harder to know whether a message, review, essay, or report was made by a person or a machine. This rapid growth of AI-generated text brings new chances to improve learning, research, and communication, but it also raises concerns about honesty, misinformation, and content quality. Recent studies show that modern models can write convincing summaries, scientific abstracts, and creative work, while detection tools often struggle to separate real and artificial content, especially when the text is slightly changed or produced with new generation methods. Research has also explored how machine written images, reviews, stories, and mathematical reasoning can affect trust in different fields. As AI systems continue to improve, many existing detection methods become less effective, highlighting the need for stronger, more reliable approaches. This work brings together key findings from recent research to help understand how AI-generated text is created, why it is difficult to identify, and why improved detection frameworks are necessary for safe and responsible use of advanced technologies. and flywheel maintain stability in self-balancing bicycles during tilting. Autonomous capability includes use of various sensors and camera to detect objects for navigation and use computer vision efficiently with the help of machine learning algorithms. The proposed design framework will help in a transformative era of innovation in urban mobility, promising safer and more sustainable ways of getting around in cities.

Keywords: AI Text Detection, Machine Learning Classifiers, Large language models (LLMs), Deep Learning Models, Bidirectional encoder representations from transformers (BERT), Support Vector Machine (NL), Text Authenticity, Academic Integrity, Plagiarism Check, Linguistic Features, Natural Language Processing (NLP)

1. Introduction

Recent progress in advanced computing and artificial intelligence has led to the development of highly powerful systems capable of producing text, images, audio, and other forms of digital media with an exceptional level of realism. These systems are primarily based on deep learning techniques and large language models that learn patterns from massive datasets, enabling them to generate content that often appears indistinguishable from human-created work. As a result, AI-generated text has become a significant topic of interest across multiple domains, including technology, education, journalism, digital media, and cybersecurity. With the rapid growth in model size and computational capacity, modern AI systems are now capable of writing detailed technical reports, generating creative narratives, summarizing lengthy documents, answering complex questions, and even producing scientific-style abstracts with coherent structure and terminology. These capabilities have greatly improved productivity by automating repetitive tasks, accelerating research workflows, and supporting content creation in both professional and academic environments. AI-based tools are increasingly being integrated into writing assistants, customer support systems, and decision-support platforms, demonstrating their broad practical usefulness. However, alongside these advantages, the widespread use of AI-generated content introduces serious ethical, social, and security-related concerns. AI-generated images and text can be used to spread misinformation, manipulate public opinion, or create misleading news content that is difficult to distinguish from authentic sources. Automated reviews and comments may artificially influence public perception, while AI-written essays and assignments raise concerns regarding academic integrity and fair assessment in educational institutions. Furthermore, malicious use of AI-generated text in phishing attacks fake research papers.

Due to these risks, it has become essential to study not only how AI-generated text is produced but also how it can be reliably detected and distinguished from human-authored content. Research in AI-generated text detection focuses on identifying linguistic patterns, statistical inconsistencies, and model-specific characteristics that reveal machine involvement. Understanding these detection techniques is critical for maintaining trust, ensuring ethical use of AI technologies, and developing policies that promote responsible deployment. As AI systems continue to evolve, effective detection and regulation will play a vital role in balancing innovation with accountability.

Despite these risks, AI also offers meaningful benefits. It can support learning, assist with research, help generate summaries, and make complex information easier to understand. To use these tools safely, clear guidelines and responsible practices are essential.

This review consolidates recent developments in AI Text Detection, The surveyed works are categorized as follows:

(i) Foundations of AI-Generated Text & Image Detection
(ii) AI-Generated Content in Scientific and Academic Contexts
(iii) Text Detection Accuracy, Limitations & Evaluation Challenges
(iv) Plagiarism, Academic Integrity & Ethical Concerns
(v) Human vs AI Writing: Distinguishability & Perception
(vi) Detection Methods & Machine Learning Approaches
(vii) Broader Impacts of LLMs in Education & Society

Another major concern is the growing impact of AI-generated content on society at large. As generative models become more sophisticated and widely accessible, the risk of misinformation spreading rapidly across digital platforms increases significantly. AI-generated text can be used to create misleading news articles, fake reviews, and manipulated narratives that are difficult for the general public to distinguish from genuine human-written content. In addition, the widespread use of AI systems raises serious concerns about personal data privacy, as these models may unintentionally expose sensitive information or be exploited to generate content that mimics individuals without consent. Ethical challenges related to fairness, transparency, and accountability are also becoming increasingly prominent, particularly when AI-generated outputs influence public opinion, academic integrity, or decision-making processes. Furthermore, many existing AI-content detection tools struggle to maintain accuracy when confronted with newly developed language models or even minor modifications in the generated text, such as paraphrasing or stylistic changes.

This limitation makes reliable detection extremely challenging in real-world scenarios. Consequently, there is a pressing need to develop more robust, adaptive, and resilient detection systems capable of operating effectively under realistic and continuously evolving conditions.

Beyond detection challenges, the unchecked proliferation of AI-generated content may also erode trust in digital communication and information sources. When users are unable to confidently verify whether content is human- or machine-generated, skepticism toward online material increases, potentially undermining the credibility of journalism, educational resources, and social media platforms. This erosion of trust can have long-term societal consequences, including reduced civic engagement and increased polarization. Moreover, AI-generated content can amplify existing biases present in training data, leading to the reinforcement of stereotypes or the marginalization of certain groups. Such outcomes highlight the importance of incorporating fairness-aware and bias-mitigation strategies into AI development and deployment.

2. Foundations of AI-generated text & image detection

Deep learning is being used to spot AI-generated images in news and journalism. As synthetic visuals spread quickly and can mislead the public, a reliable method is needed to separate real images from artificial ones. A CNN model trained on both real and AI-created images including those from modern diffusion models shows strong accuracy and learns useful patterns quickly. It also performs better and faster than models like ResNet50 and InceptionV3, making it practical for real-time news environments. With further testing on broader datasets, it can become an effective tool against misinformation [1].

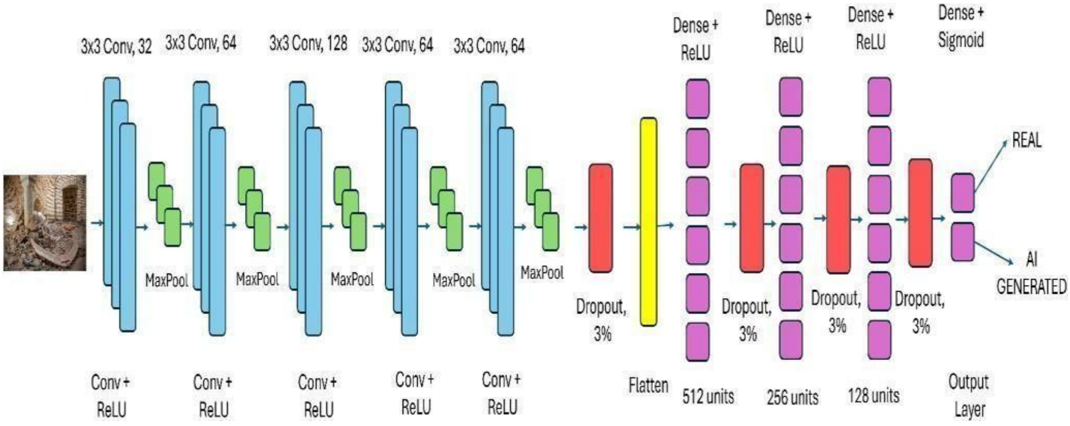


Figure1: Proposed CNN Architecture [1]

A new method called DetectGPT helps identify text produced by large language models. It works by examining how the model's probability surface behaves and uses slight rewrites of the text to compare scores. This approach requires no extra training or labeled datasets and outperforms earlier zero-shot detection methods. It also opens new research opportunities, such as combining watermarking with probability-based detection and extending the idea to other generative fields like audio or images [2]. Growing advances in natural language generation make it difficult to tell human and machine-written text apart. A broad survey outlines the key risks, including misinformation and automated content manipulation, and reviews many detection methods used today. The study stresses that detection systems must be fair, robust, and transparent. Current approaches often fail under real-world conditions, adversarial threats, or unfamiliar model architectures. Addressing these limitations will require joint efforts from researchers, cybersecurity specialists, and policymakers to create more dependable detection systems [3].

Recent progress in generative models has also increased concerns around deepfake text. An evaluation using text from several online Transformer-based tools shows that many existing detection systems fail when tested on real-world examples. Their performance drops sharply when exposed to simple adversarial attacks, revealing weaknesses that controlled tests often hide. The study also proposes a new attack method that works without direct access to the target model [4].

3. AI-Generated Content in Scientific and Academic Contexts

Researchers studied how to tell human-written scientific abstracts from those created by GPT- 3, as AI-generated content is becoming harder to notice. They used several machine- learning models, from basic text-feature methods to advanced deep-learning approaches, to check how well different techniques could identify synthetic writing. Results showed that many models can already detect GPT-3-generated abstracts with good accuracy. The study also suggests building larger and more varied datasets in the future to test performance across different subjects, languages, and writing styles, helping improve the responsible use of AI- generated text [5].

Researchers compared real medical journal abstracts with abstracts generated by ChatGPT using only paper titles. An AI-detection tool labeled most AI-generated abstracts as fake with very strong confidence, while real abstracts scored very low. Human reviewers were able to identify many of the AI-generated abstracts but sometimes mistakenly labeled real ones as machine-written. Reviewers found that the AI versions sounded general and repetitive, even though they appeared realistic. The study shows that ChatGPT can create convincing text, but the information is invented, so detection tools and clear guidelines are important for maintaining scientific quality [6].

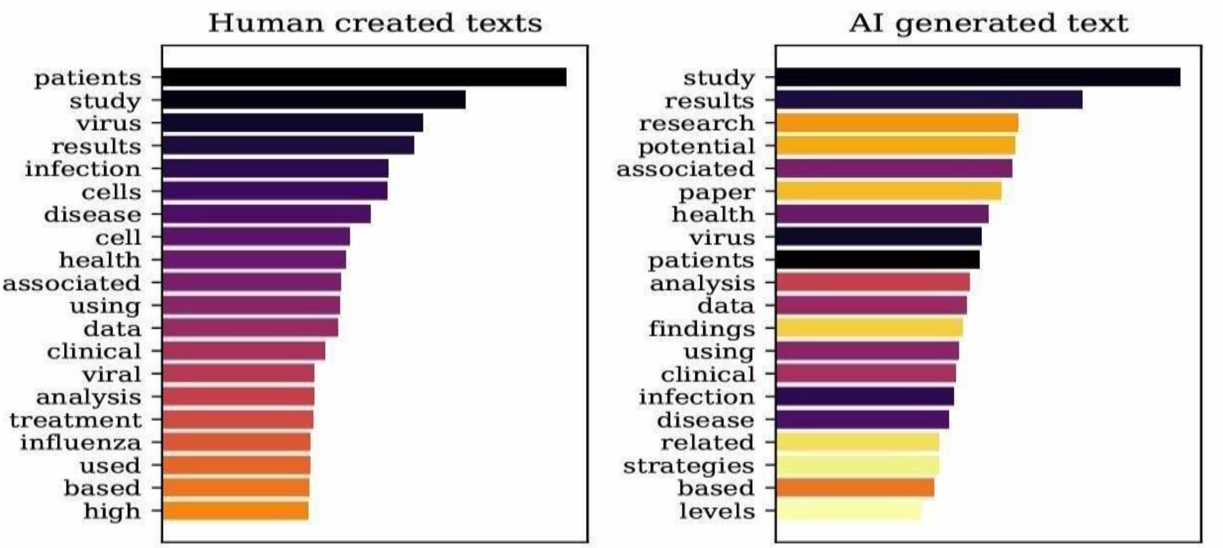
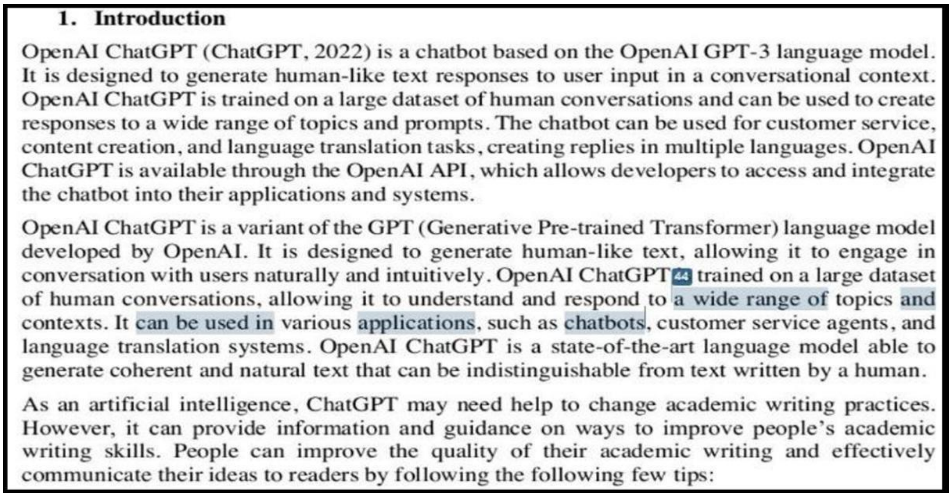


Figure 2: Most frequent word appearances in both the human-created texts(left) and the AI- generated texts(right) [6]

A study explored whether ChatGPT could help create literature reviews by paraphrasing abstracts about Digital Twin technology in healthcare. While the generated summaries looked clear and well-structured, plagiarism-checking tools showed that the paraphrased text still had high similarity to the originals [7].



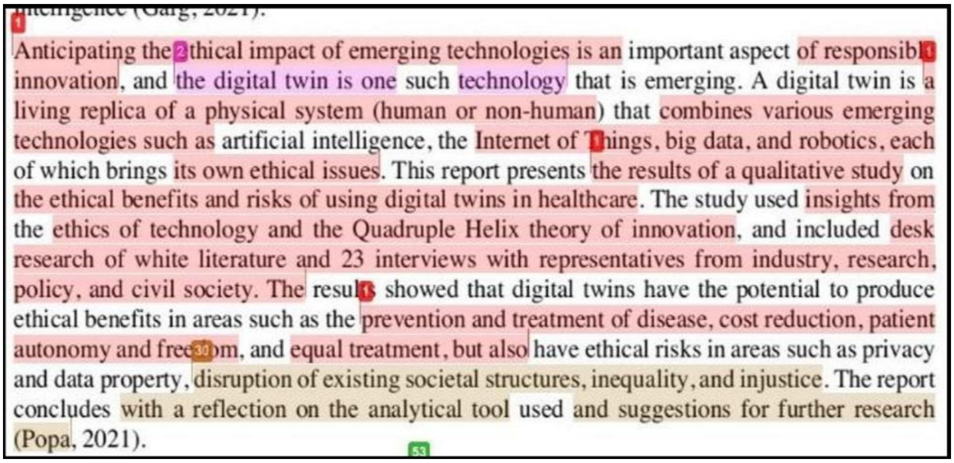
1. Introduction

OpenAI ChatGPT (ChatGPT, 2022) is a chatbot based on the OpenAI GPT-3 language model. It is designed to generate human-like text responses to user input in a conversational context. OpenAI ChatGPT is trained on a large dataset of human conversations and can be used to create responses to a wide range of topics and prompts. The chatbot can be used for customer service, content creation, and language translation tasks, creating replies in multiple languages. OpenAI ChatGPT is available through the OpenAI API, which allows developers to access and integrate the chatbot into their applications and systems.

OpenAI ChatGPT is a variant of the GPT (Generative Pre-trained Transformer) language model developed by OpenAI. It is designed to generate human-like text, allowing it to engage in conversation with users naturally and intuitively. OpenAI ChatGPT trained on a large dataset of human conversations, allowing it to understand and respond to a wide range of topics and contexts. It can be used in various applications, such as chatbots, customer service agents, and language translation systems. OpenAI ChatGPT is a state-of-the-art language model able to generate coherent and natural text that can be indistinguishable from text written by a human.

As an artificial intelligence, ChatGPT may need help to change academic writing practices. However, it can provide information and guidance on ways to improve people's academic writing skills. People can improve the quality of their academic writing and effectively communicate their ideas to readers by following the following few tips:

Figure 3 : Plagiarism Tool Match Screenshot for The Authors' writings [7]



intelligence (Garg, 2021).

Anticipating the ethical impact of emerging technologies is an important aspect of responsible innovation, and the digital twin is one such technology that is emerging. A digital twin is a living replica of a physical system (human or non-human) that combines various emerging technologies such as artificial intelligence, the Internet of Things, big data, and robotics, each of which brings its own ethical issues. This report presents the results of a qualitative study on the ethical benefits and risks of using digital twins in healthcare. The study used insights from the ethics of technology and the Quadruple Helix theory of innovation, and included desk research of white literature and 23 interviews with representatives from industry, research, policy, and civil society. The results showed that digital twins have the potential to produce ethical benefits in areas such as the prevention and treatment of disease, cost reduction, patient autonomy and freedom, and equal treatment, but also have ethical risks in areas such as privacy and data property, disruption of existing societal structures, inequality, and injustice. The report concludes with a reflection on the analytical tool used and suggestions for further research (Popa, 2021).

Figure 4 : Plagiarism Tool Match Screenshots for ChatGPT Paraphrased Abstracts [7]

In contrast, parts written by the authors had very low similarity. This shows that AI tools can speed up the gathering and summarizing of information, but may not produce fully original paraphrased content. As academic work evolves, AI may support efficiency, but researchers must still ensure originality and accuracy [7].

A survey of about 1,100 first-semester engineering students done that although most believed they understood academic integrity, many struggled to correctly identify plagiarism. Students often misunderstood rules about quoting and paraphrasing, even though most reported receiving previous training on academic misconduct. These results suggest that mistakes may come from confusion rather than intentional cheating. The study stresses the need for clearer teaching materials, stronger examples, and better assessment tools to help students understand proper citation and avoid unintentional plagiarism [8].

4. Text Detection Accuracy, Limitations & Evaluation Challenges

Modern large language models can generate text that closely resembles human writing, making it challenging for humans and automated systems to differentiate between AI- generated and human-written content. Studies show that humans often rely on meaning-based errors to detect AI text, whereas automated detectors focus on statistical patterns introduced by decoding strategies like top-k, nucleus, or random sampling. While humans are fooled over 30% of the time even on longer passages, automated detectors are going to achieve higher accuracy, especially when trained across multiple sampling strategies and given longer text [9].

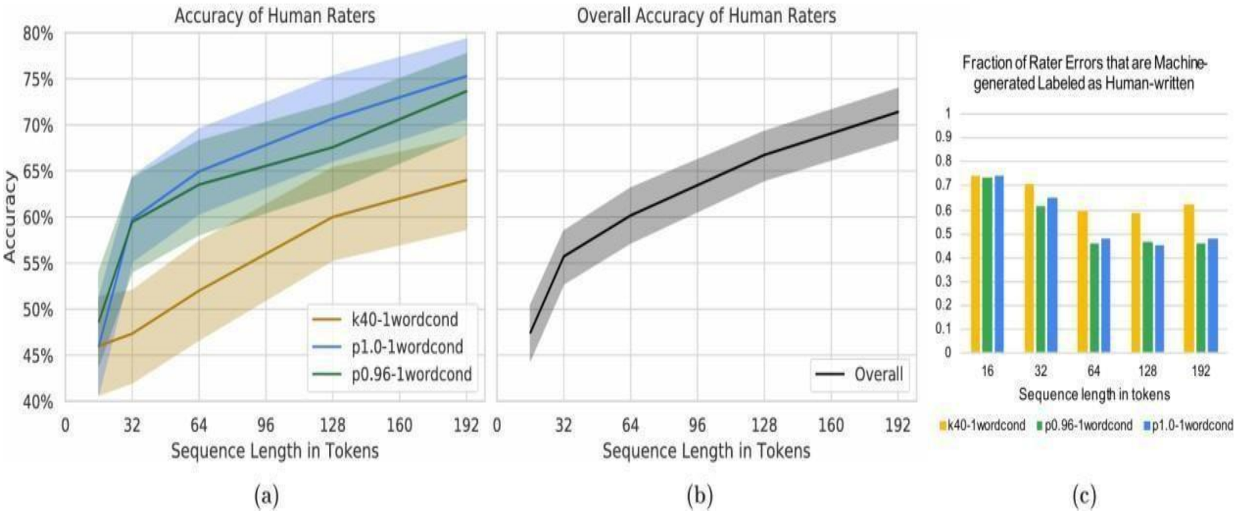


Figure 6 : (a) and (b) show human rater accuracy of correctly identifying an excerpt as human-written or machine written, shown with 80% confidence internals, in (a), broken up by decoding strategy and in (b), overall. Accuracy increases as raters observe more tokens. (c) shows that for short excerpts, most rater mistakes are them incorrectly thinking machine-generated text is human written. The two errors types become more balanced at longer lengths [9]

The widespread adoption of models like ChatGPT introduces additional evaluation challenges due to their closed-source nature, continuous updates, and potential data contamination. Ensuring fair testing is difficult because these models may have been exposed to test datasets during training. Moreover, LLMs are often trained on massive digital corpora containing sensitive or proprietary material, and current techniques to prevent misuse or data leakage are limited [10].

In academic contexts, reliance on plagiarism-detection tools such as Turnitin and MyDropBox is often misplaced. Studies reveal that these tools frequently fail to detect copied content from journals or paywalled sources, performing reliably only on openly accessible web material. This overestimation of effectiveness can mislead institutions into assuming submitted work is plagiarism-free when it may not be [11].

Overall, the rapid proliferation of AI-generated text has prompted the development of numerous detection methods, but the field still lacks standardized evaluation metrics and comprehensive understanding of model limitations. Persistent challenges include detecting outputs from increasingly sophisticated models, handling open-source variants, and addressing adversarial risks. Future work requires robust measurement frameworks, adaptive detection strategies, and governance approaches that ensure responsible use of language generation technologies [12].

5. Plagiarism, Academic Integrity & Ethical Concerns

Scientific integrity requires researchers to follow clear methods, rely on real evidence, obtain informed consent, and avoid reusing published text without permission, as plagiarism undermines trust and can produce false findings. Universities and research organizations enforce ethical standards, often guided by bodies such as EASE, WAME, and COPE, while journals may retract plagiarized work and blacklist authors. In medicine and other fields, proper citation, access to reliable databases, and adherence to ethical practices are crucial for maintaining credibility, even as pressures to publish increase the risk of misconduct [13].

Detecting text generated by large language models has become a critical area of study. Methods include machine-learning classifiers (Random Forest, SVM, XGBoost) trained on handcrafted linguistic, stylistic, and n-gram features, achieving near-perfect performance in distinguishing human from AI-generated text. For scenarios without human reference text, techniques such as topic extraction and cosine similarity comparisons with LLM outputs also provide effective discrimination. Key predictive features include readability scores, word density, punctuation, error patterns, and title-word counts. These approaches can be further strengthened by reverse- engineering LLM behaviors and incorporating insights from AI models themselves [14].

Research on dishonest behavior in children shows that cognitive ability, socioeconomic background, and altruistic tendencies influence the likelihood of cheating. Higher-IQ children and those from advantaged households tend to cheat more, while altruistic children are less likely to cheat when rewards are introduced. Incentive structures in school- like settings have limited impact, suggesting that early patterns of dishonesty are shaped more by internal and environmental factors than by immediate external rewards. Understanding these early behaviors is crucial, as they may develop into long-term habits [15].

6. Human Vs AI Writing: Distinguishability & Perception

The growing use of AI for text generation has prompted research into distinguishing human- written from AI-generated content. Methods leveraging transfer learning on datasets of real and AI-generated book reviews, such as those created with Vicuna, achieve high accuracy (96.86%), though subtle word choices can still make human and AI text difficult to differentiate. These approaches can be extended to other text types, languages, and formats to strengthen detection and preserve authenticity [16].

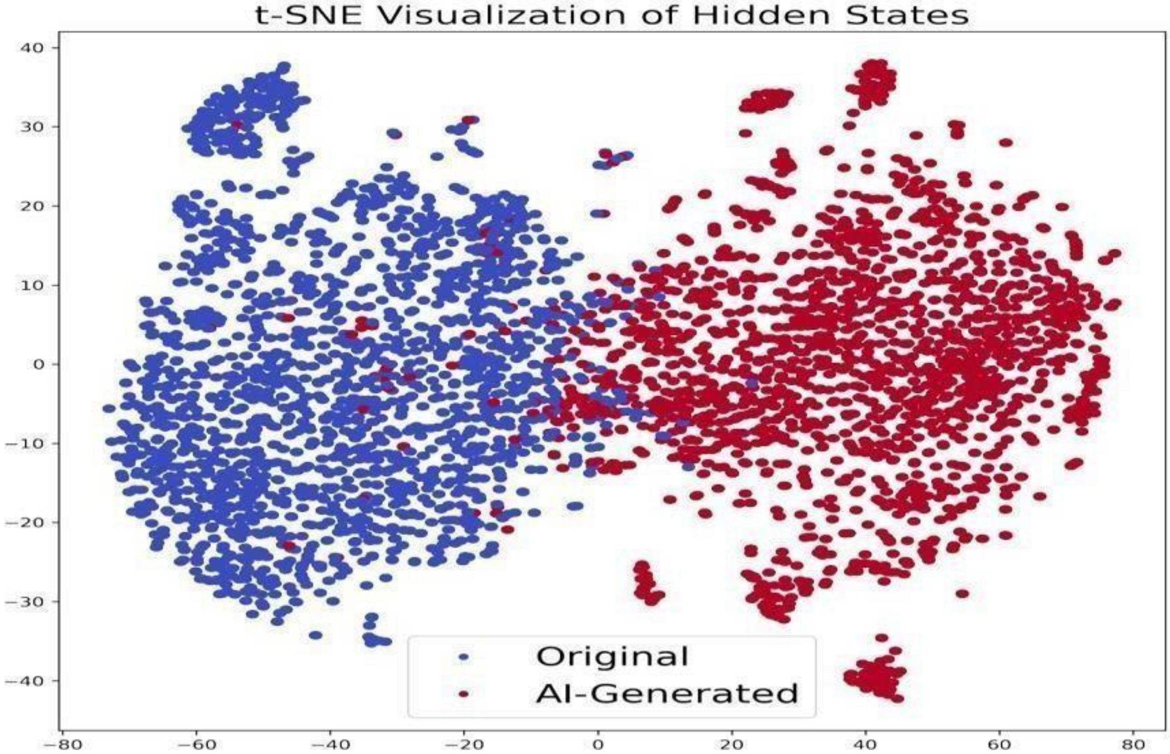
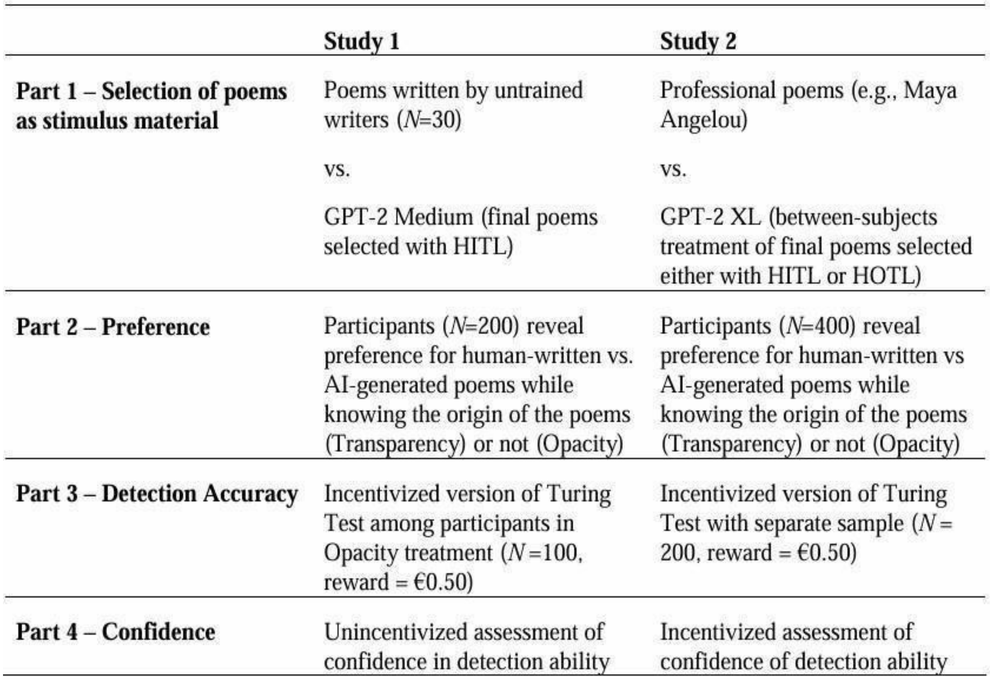


Figure 7: t-SNE Visualization of Hidden State Representation on the evaluation data. The blue points represent the original reviews and the red are the AI generated text [16]

Studies comparing large language models like ChatGPT and Vicuna to human language processing show that these models replicate many humanlike behaviors. ChatGPT, in particular, mirrors human performance in ten out of twelve cognitive experiments, including word meaning inference, sentence structure repetition, and context-sensitive word choice. Vicuna demonstrates similar but slightly fewer humanlike patterns. Both models, however, differ from humans in areas such as word-length preference and resolving certain syntactic ambiguities, indicating that LLMs are not perfect analogy of human cognition [17]. Evaluations of AI-generated essays reveal that models consistently produce higher-scoring argumentative writing compared to human students. Their essays exhibit richer vocabulary and more nominalizations but fewer discourse markers and epistemic cues, reflecting a stylistic rather than cognitive alignment with human student writing. These findings suggest that conventional assessment practices may no longer accurately measure student ability and that educational strategies should evolve to integrate AI as a tool for higher-level reasoning and critical engagement [18]. Experiments comparing human- and AI-generated poetry demonstrate that people often cannot reliably distinguish AI outputs from human work, especially when AI outputs are curated [19].

Table 1 : Overview of two studies that each contain four parts [19]



	Study 1	**Study 2**
Part 1 – Selection of poems as stimulus material	Poems written by untrained writers (*N*=30) vs. GPT-2 Medium (final poems selected with HITL)	Professional poems (e.g., Maya Angelou) vs. GPT-2 XL (between-subjects treatment of final poems selected either with HITL or HOTL)
Part 2 – Preference	Participants (*N*=200) reveal preference for human-written vs. AI-generated poems while knowing the origin of the poems (Transparency) or not (Opacity)	Participants (*N*=400) reveal preference for human-written vs AI-generated poems while knowing the origin of the poems (Transparency) or not (Opacity)
Part 3 – Detection Accuracy	Incentivized version of Turing Test among participants in Opacity treatment (*N* =100, reward = €0.50)	Incentivized version of Turing Test with separate sample (*N* = 200, reward = €0.50)
Part 4 – Confidence	Unincentivized assessment of confidence in detection ability	Incentivized assessment of confidence of detection ability

While participants show a mild preference for human-authored poems, the results highlight AI's growing ability to mimic human creativity. These behavioral insights are essential for developing guidelines on transparency, disclosure, and responsible use of AI in creative domains [19].

7. Detection Methods & Machine Learning Approaches

One study introduces a system for separating ChatGPT-generated text from human text using 11 different models and a dataset of 10,000 samples. The best model achieved 77% accuracy when tested on GPT-3.5 outputs, emphasizing the difficulty of detection as models continue to improve [20]. Another work evaluates classical machine-learning methods alongside a BERT- based deep-learning model. BERT significantly outperformed others, reaching 93% accuracy, demonstrating its superior ability to capture contextual and stylistic cues in human vs. AI writing. The study also highlights ethical, transparency, and societal concerns related to widespread AI text generation [21].

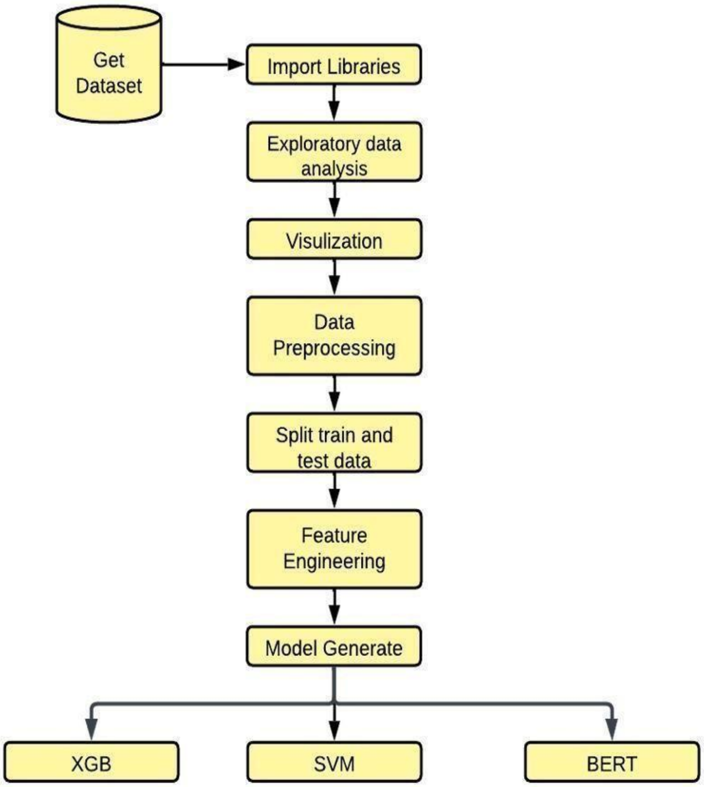


Figure 8 : The proposed methodology of the work [21]

The figure 8 presents a word cloud generated from the dataset's text. In this type of visualization, words appear in larger or smaller sizes based on how often they occur, allowing the most common and meaningful terms to stand out immediately. By turning extensive text into an easy-to-scan graphic, a word cloud helps reveal dominant themes and recurring ideas at a glance. Because of this, it is widely used in areas such as data analysis, market research, and text exploration to highlight patterns and key topics within large collections of documents [21].

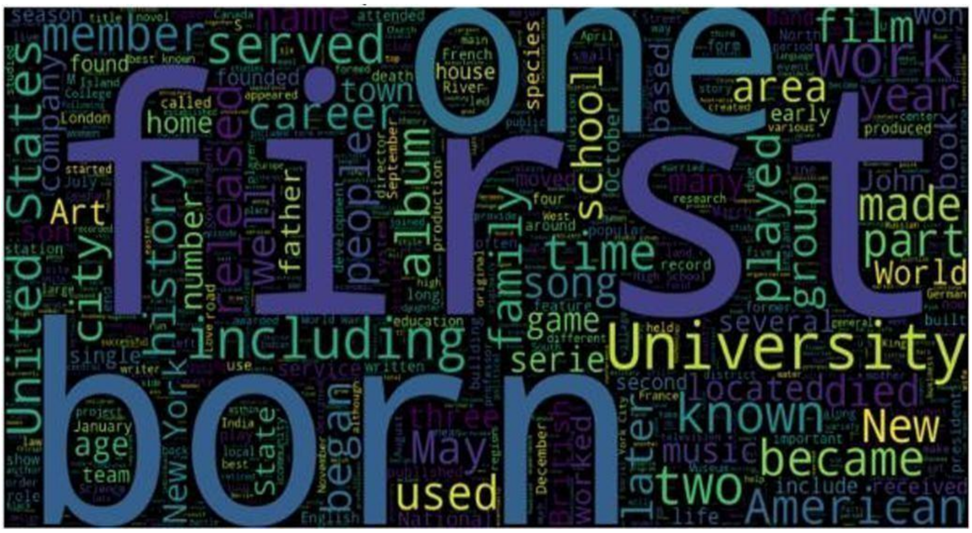


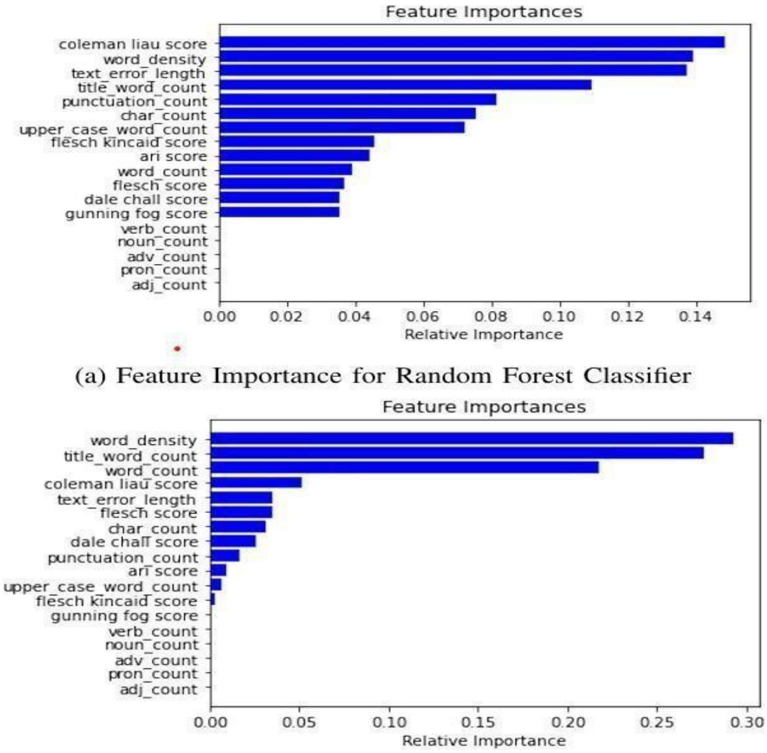
Figure 9: Word cloud of text dataset [21]

The results shown in the table 2 BERT clearly outperformed the other models in detecting AI-generated text, achieving 93% accuracy. XGBoost reached 84% and SVM 81%, indicating solid but comparatively lower performance. BERT's advantage comes from its ability to understand word meaning in context, allowing it to capture subtle linguistic patterns that distinguish human writing from machine-produced content. While XGBoost and SVM provide reliable results, they are less effective at handling the deeper language nuances that BERT can recognize. Overall, the findings highlight BERT as the most capable model for this classification task [21].

Table 2: Performances of Different Classifiers [21]

Algorithm Name	Class	Precision	Recall	F1 Score	Accuracy
XGB Classifier	0	0.86	0.82	0.84	0.84
	1	0.83	0.87	0.85	
SVM	0	0.79	0.83	0.81	0.81
	1	0.82	0.78	0.80	

A third study explores two detection strategies: a machine-learning feature-based approach and a text-similarity approach. Using handcrafted linguistic, stylistic, and topic- based features, models such as Random Forest and XGBoost achieved near-perfect F1 scores (up to 0.9993) on datasets combining human text with ChatGPT-generated content [22].



(a) Feature Importance for Random Forest Classifier

(b) Feature Importance for XGB Classifier

Figure 10 : Feature Importance Classifiers [22]

The text-similarity method works even when no human reference text is available by comparing generated topics and cosine similarity scores. Feature-importance analysis reveals that metrics like the Coleman–Liau score, word density, punctuation patterns, and error-related features are the strongest predictors for detection. The authors suggest that future work could leverage LLMs themselves to build stronger detection systems [22].

8. Broader Impacts of LLMS In Education & Society

Recent research highlights the expanding influence of AI-generated content and large language models across multiple domains, while also addressing the technical, ethical, and evaluative challenges that accompany their growth. A broad survey of AIGC systems outlines how models like ChatGPT generate realistic text and images, noting the security, privacy, ethical, and legal concerns associated with their widespread use [23].

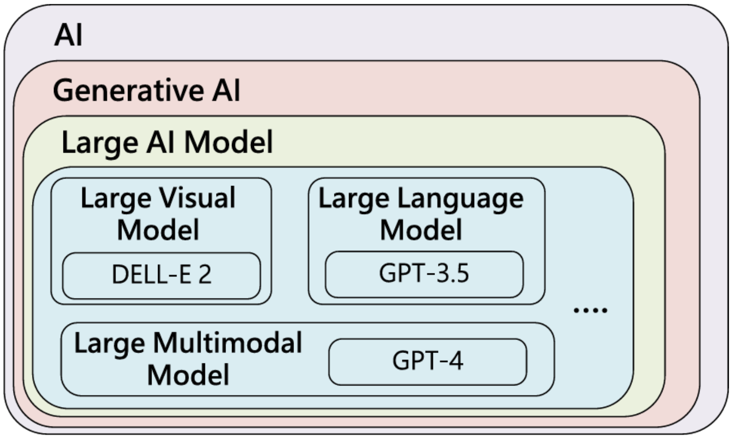


Figure 11: Relation between existing representative large AI models and AIGC. Generative AI algorithms are a class of AI algorithms that create new content in various forms (e.g., images, text, and music) by learning underlying patterns from training data. AIGC encompasses a broader scope and includes not only generative AI algorithms but also other AI techniques such as natural language processing and computer vision. A large AI model refers to any neural network architecture that has large number of parameters, such as large visual model (LVM), large language model and large multimodal model [23]

Issues such as jailbreak attacks, deepfakes, data leaks, and biased or harmful outputs remain difficult to control. Current mitigation strategies—including watermarking and detection methods—offer partial solutions but are not yet sufficient to fully govern the rapid expansion of AI-generated media. The study emphasizes the need for future AIGC systems that are more transparent, accountable, environmentally efficient, and resistant to adversarial attacks [23].

In education, LLMs are transforming both instructional support and assessment by enhancing reading, writing, speaking, and tutoring systems. The integration of LLMs into established NLP-based educational technologies has led to more adaptive and inclusive learning experiences. However, major challenges persist, especially regarding limited training data, the need for reliable evaluation frameworks, and ethical concerns involving fairness and transparency. Collaboration across researchers, educators, and interdisciplinary specialists is viewed as essential for developing effective assistive tools and assessment models for classrooms of the future [24]. Other ll,work examines the mathematical reasoning capabilities of modern language models, particularly their arithmetic skills, which underlie more complex chain-of-thought reasoning. A new benchmark, MATH 401, evaluates arithmetic proficiency among models such as GPT-4, ChatGPT, Galactica, InstructGPT, and LLaMA. Findings show that factors like training data, tokenization, model size, and prompting strategies significantly affect arithmetic accuracy [25].

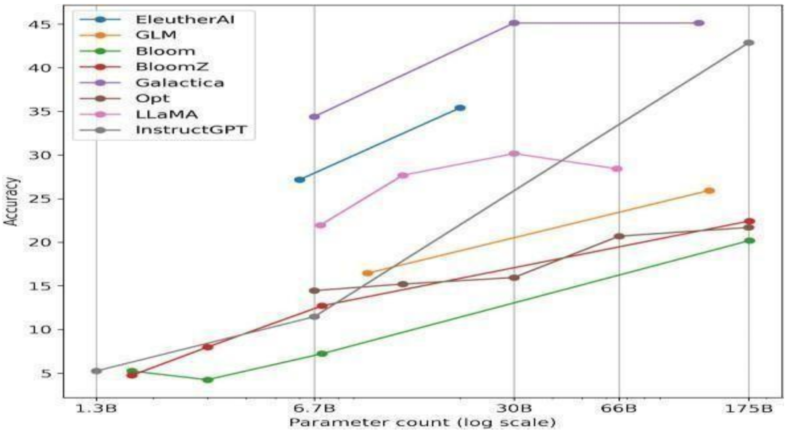


Figure 12 : Performances of MATH 401 on LLMs with different sizes. We do not know the parameter count of ChatGPT. We list InstructGPT results with SFT setting (text-davinci-002) only for a fair comparison [25]

ChatGPT performs especially well, though the reasons behind this strength remain partly unexplained. The authors propose extending this type of evaluation into broader mathematical domains including algebra, geometry, and symbolic reasoning to better understand LLM mathematical performance [25].

In Table 3 The confusion matrix makes it clear that the system usually misses the correct score by only one level, which is not surprising because the annotation process itself allowed humans to disagree by one point without requiring a third reviewer. The biggest difficulties appear at the extreme ends of the scoring scale scores 0, 1, and 4 where the system either over- or underestimates more often. A closer look at the essays that humans rated as 0 shows why. Only a small number of these were also scored 0 by the system, and those tended to be extremely short and clearly non-narrative. Most of the remaining essays were long, well- constructed pieces that did not follow the required narrative format. Because the system focuses mainly on surface qualities such as length and fluency, it gave these off-purpose essays higher marks than a human reader would [26].

The same pattern shows up for essays that received a 1 from the annotators. Short, weak responses were pushed down to 0 by the system, while longer essays especially those that drifted toward expository writing were often scored too high. The machine seemed to reward length more than content or purpose, which led to a noticeable number of over-scored essays. At the upper end of the scale, essays that humans judged as 4 were sometimes given lower scores by the machine, especially when they were relatively short. Taken together, the analysis shows two main issues: the system struggles with off-purpose, non-narrative responses, and it is overly sensitive to essay length. Addressing both problems possibly by adding a narrative-vs-non-narrative classifier and reducing reliance on length would likely improve scoring accuracy [26].

Table 3 : Human machine confusion matrix for Development traits scores [26]

Human	Machine					
	0	1	2	3	4	Total
0	8	9	18	5	0	40
1	8	28	43	5	0	84
2	1	8	159	101	1	270
3	0	0	83	205	31	319
4	0	0	9	125	95	229

Narrative focused linguistic features were then used to train models that can predict writing quality. Results indicate that narrative-specific features outperform general writing features for certain storytelling traits. While creativity and open-ended story structures make automated scoring difficult, the findings suggest strong potential. Future improvements will require deeper modeling of narrative elements such as plot, character development, point of view, and how these components interact within a story [26].

Conclusion

AI-generated text is now common in many areas of life, and it is getting harder to tell it apart from writing done by people. The studies show that these tools can create clear and believable text, which can be helpful but also risky. Because of this, many groups are trying to build better ways to check if something was written by a machine or a human. Some detection methods work well in controlled tests, but many fail when the text is changed even slightly or when newer models are used. This makes it clear that current tools are not enough on their own. There is a strong need for better systems that can handle different writing styles, topics, and real-world situations. To move forward, both technical improvements and responsible use are important. Developers, teachers, editors, and policy makers need to work together to make sure AI-generated text is used in safe and honest ways. With the right balance, society can benefit from these new technologies while also reducing the risks that come with them.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article.

Funding

The author received no financial support for the research, authorship and publication of this article.

References

1. Tarun Jagadish, Graceline Jasmine S “Detection of AI-Generated Image Content in News and Journalism” 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-

7024-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCCNT61001.2024.10724589

2. Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D. Manning, Chelsea Finn "DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature" arXiv:2301.11305v2 [cs.CL] 23 Jul 2023
3. EVAN N. CROTHERS, NATHALIE JAPKOWICZ, ANDHERNAL. VIKTOR "Machine-Generated Text: A Comprehensive Survey of Threat Models and Detection Methods "Digital Object Identifier 10.1109/ACCESS.2023.3294090 version 18 July 2023.
4. Jiameng Pu, Zain Sarwar, Sifat Muhammad Abdullah, Abdullah Rehman, Yoonjin KimParantapa Bhattacharya, Mobin Javed, Bimal Viswanath "Deepfake Text Detection: Limitations and Opportunities" arXiv:2210.09421v1 [cs.CR] 17 Oct 2022
5. Panagiotis C. Theocharopoulos, Panagiotis Anagnostou, Anastasia Tsoukala, Spiros V. Georgakopoulos, Sotiris K. Tasoulis, Vassilis P. Plagianakos "Detection of Fake Generated Scientific Abstracts" arXiv:2304.06148v1 [cs.CL] 12 Apr 2023
6. Catherine A. Gao, Frederick M. Howard, Alexander T. Pearson, Nikolay S. Markov, Emma C. Dyer, Siddhi Ramesh, Yuan Luo "Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers"
7. Ömer AYDIN, Enis Karaarslan "OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare" Published in Social Science Research…2022 (pp. 22-31)
8. Dr. Susan L. Murray, Dr. Amber M. Henslee, Dr. Douglas K. Ludlow "Engineering Students Understanding of Plagiarism" Paper ID #11591 122 nd ASEE Annual conference & Exposition June 14-17, 2015
9. Daphne Ippolito, Daniel Duckworth, Chris Callison-Burch, Douglas Eck "Automatic Detection of Generated Text is Easiest when Humans are Fooled" arXiv:1911.00650v2 [cs.CL] 7 May 2020
10. Rachith Aiyappa, a Jisun An, a Haewoon Kwak, Yong-Yeol Ahna "Can we trust the evaluation on ChatGPT?" arXiv:2303.12767v2 [cs.CL] 22 Aug 2024
11. Rebecca L. Fiedler, Cem Kaner Plagiarism "Plagiarism Detection Services: How Well Do They Actually Perform?" IEEE TECHNOLOGY AND SOCIETY MAGAZINE | WINTER 2010 1932-4529/10/$26.00©2010IEEE
12. Ruixiang Tang, Yu-Neng Chuang, Xia Hu "The Science of Detecting LLM- Generated Texts" arXiv:2303.07205v3 [cs.CL] 2 Jun 2023
13. Izet Masic "Plagiarism in Scientific Research and Publications and How to Prevent It" DOI: 10.5455/msm.2014.26.141-146 Mater Sociomed 2014 Apr; 26
14. Mohammad Khalil, Erkan Er "Will ChatGPT get you caught? Rethinking ofPlagiarism Detection" conference paper First Online:09 June 2023
15. Sule Alan, Seda Ertac, Mert Gümren "Cheating and Incentives in a Performance Context:Evidence from a Field Experiment on Children" https://doi.org/10.1016/j.jebo.2019.03.015 Volume 179, November 2020, Pages 681-701
16. Panagiotis C. Theocharopoulos, Spiros V. Georgakopoulos, Sotiris K. Tasoulis, Vassilis P. Plagianakos "Who Writes the Review, Human or AI?" arXiv:2405.20285v1 [cs.CL] 30 May 2024
17. Zhenguang G. Cai, Xufeng Duan, David A. Haslett, Shuqi Wang, Martin J. Pickering "Do large language models resemble humans in language use?" last revised 26 Mar 2024 (this version, v2)
18. Steffen Herbold, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva,Alexander Trautsch "AI, write an essay for me: A large-scale comparison of human-written versus ChatGPT- generated essays" arXiv:2304.14276v1 [cs.CL] 24 Apr 2023
19. Nils Köbis, Luca D. Mossink "Artificial Intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry" https://doi.org/10.1016/j.chb.2020.106553 Volume 114, January 2021
20. Niful Islam, Debopom Sutradhar, Humaira Noor, Jarin Tasnim Raya, Monowara Tabassum Maisha, Dewan Md Farid "Distinguishing Human Generated Text From ChatGPT Generated Text Using Machine Learning" arXiv:2306.01761v1 [cs.CL] 26 May 2023
21. Nuzhat Noor Islam Prova "Detecting AI Generated Text Based on NLP and Machine Learning Approaches" [v1] Mon, 15 Apr 2024
22. Trung T. Nguyen, Amartya Hatua, Andrew H. Sung "How to Detect AI-Generated Texts?" 2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) 979-8-3503-0413-8/23/$31.00 ©2023 IEEE DOI: 10.1109/UEMCON59035.2023.10316132

23. Yuntao Wang, Yanghe Pan, Miaoyan, Zhou Su (Senior Member, IEEE), Tom H. Luan "A Survey on ChatGPT: AI–Generated Contents, Challenges, and Solutions" Digital Object Identifier 10.1109/OJCS.2023.3300321
24. Sowmya Vajjala, Bashar Alhafni, Stefano Bannò, Kaushal Kumar Maurya, Ekaterina Kochmar "Opportunities and Challenges of LLMs in Education: An NLP Perspective" arXiv:2507.22753v1 [cs.CL] 30 Jul 2025
25. Zheng Yuan, Hongyi Yuan, Chuanqi Tan, Wei Wang, Songfang Huang "How well do Large Language Models perform in Arithmetic tasks?" arXiv:2304.02015v1 [cs.CL] 16 Mar 2023
26. Swapna Somasundaran, Michael Flor, Martin Chodorow, Hillary Molloy, Binod Gyawali, Laura McCulla Towards "Evaluating Narrative Quality In Student Writing Transactions of the Association for Computational Linguistics" vol. 6, pp. 91–106, 2018. Action Editor: Alexander Clark. Submission batch: 7/2017; Revision batch: 10/2017; Published 2/2018