BULETINUL INSTITUTULUI POLITEHNIC DIN IAŞI
Publicat de
Universitatea Tehnică „Gheorghe Asachi” din Iaşi
Volumul 71 (75), Numărul 3, 2025
Secţia
CONSTRUCŢII DE MAŞINI
DOI: 10.2478/bipcm-2025-0021

THE LIMITATIONS OF EXISTING SOFTWARE IN IDENTIFYING CONTENT CREATED WITH THE HELP OF AI

BY

EMANUEL MIHĂLUȚE[1] and NICOLAE-RĂZVAN MITITELU[2,*]

[1]Mechanical Department at Continental Automotive Romania SRL
[2]“Gheorghe Asachi” Technical University of Iași, Department of Machine Manufacturing Technology, Iași, Romania

Received: June 12, 2025
Accepted for publication: June 19, 2025

Abstract. This paper analyzes the discrepancies between the historical reality of certain poems written decades or even centuries ago, and the results provided by modern programs designed to detect AI (Artificial Intelligence) generated content. These discrepancies are important because they can lead to errors in the authentic evaluation of literary texts affecting both the recognition of the cultural value of these works and the editorial process. Such a phenomenon can contribute to a distorted perception of classical authors and a lack of confidence in the validity of literature transmitted over the centuries.

In the academic environment, some classical texts analyzed using such AI detectors have been erroneously marked as artificially generated, which has led to unnecessary re-evaluations of their authenticity or even the rejection of academic work. Through a case study on several classical and contemporary poems scanned with such a tool, we observe how AI detectors can incorrectly classify authentic texts as AI-generated. These errors raise issues concerning the use of such instruments in evaluating the originality of literary content especially in academic and editorial contexts.

*Corresponding author; *e-mail*: nicolae-razvan.mititelu@student.tuiasi.ro
© 2025 Emanuel Mihălțe and Nicolae-Răzvan Mititelu
This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

Keywords: AI misclassification, Stylometric analysis, AI detector, Academic integrity.

1. Introduction

Over the last few years, AI programs employed to detect machine-created content have begun becoming common in scholarly, journalistic, and educational settings (Floridi, 2019). Such tools aim to detect if texts have been created by humans or generated using sophisticated language models such as GPT (Brown *et al*., 2020). Their precision, though, has been disputed. Surprisingly, some famous literary compositions composed many years ago before the advent of AI are frequently incorrectly identified as artificial, and this has raised questions regarding how such tools evaluate and authenticate literary tradition (Juola, 2008).

The following study considers these discrepancies by using a case study of 20 poems both traditional and contemporary by submitting them to an AI detection tool. The outcome reveals that the system has a very high chance of assigning AI authorship to conventionally renowned pieces such as Thomas Moore's The Last Rose of Summer (100% chance of AI authorship), Mihai Eminescu's Luceafărul (96%), and the English translation of Victor Hugo's Tomorrow, at Dawn (95%). On the other hand, the original works in their original languages e.g., Bécquer's Rima LIII (26%) and Demain, dès l'aube (68%) were assessed at much lower ratings. What that means is that AI detection outcomes can be influenced by the translation process in important ways (Hoover, 2018).

These facts point to a more extensive issue. The majority of the AI detectors are trained to identify modern corpora and therefore are not as effective in identifying ancient patterns of style. Therefore, ancient texts rich in double layers of grammar, old-fashioned phrases, and veiled metaphors are easily misidentified as AI. The special rhythm and recurrence that are usually present in poetry can also mislead the algorithms in labeling them as AI incorrectly in (Kestemont *et al*., 2019).

Religious and philosophical texts are also affected to the same extent. For instance, a 70% AI score was assigned to a sample from The Gospel of Matthew, whereas the modern religious book Arsenie Boca's The Call of Jesus received an 86% rating. On the other end of the scale, the philosophical book The Twilight of Thought by Emil Cioran was properly identified to have been authored by humans (0% AI). These variations imply that AI detectors rely more on formal elements than on any real understanding of meaning or literary value (Craig and Kinney, 2009).

These findings are reasons for alarm in the application of these systems in authenticating literary works. If a classical work is inappropriately labeled as AI generated, this would trigger unjustified doubts about its authenticity,

publishing delays, or even damage to an author's reputation. In academic circles, editors may not readily accept or review a paper based on a false positive. The problem is even worse with translations, since stylistic variation can deceive detection systems and increase the risk of misclassification. For example, the English version of Hugo's Tomorrow, at Dawn scored 95%, while the original French version could only attain 68% an appreciable difference largely because of language variability brought about by translation.

Yet, another significant point is the ethical and methodological problems in excess dependence on these instruments (Eubanks, 2018). When AI detectors are the standard in gauging literary authenticity, questionable choices are bound to ensue especially when software disregards the stylistic and historical richness of literary work (Mittelstadt *et al*., 2016). There already are cases where instructors or editors rejected authentic texts based on computerized opinions. To weigh only AI in these situations may leave significant works beyond academic or editorial realms (Noble, 2018).

This underscores the necessity of enhancing validation processes by employing algorithmic analysis in combination with human knowledge. Hybrid systems with the capacity to analyze stylistic trends of a text against corroborated literary records are a probable solution. These systems are already in practice in research projects such as GLARE (Great Library Automatic Recognition Engine), with AI assist and philological insight to determine manuscript authenticity (Andrews, 2019). Also, the HathiTrust Research Center uses large literary corpora-trained neural networks to identify stylistic deviations in previous works (Organisciak *et al*., 2017). These efforts offer a more accurate and context-aware alternative to traditional detectors.

Our research aims at unearthing the causes of these inaccuracies and proposing improvements. Most specifically, we are interested in the translation function, stylistic shift, and training materials used by AI models. At the same time, we advocate complementary strategies such as hybrid stylometric techniques and contextual authentication based on historical evidence to improve the credibility of AI assessment. In this way, we not only expose current limitations but also highlight the necessity of a more interdisciplinary method, one that unites linguistic knowledge, human critical reasoning, and modern developments in natural language processing (Moretti, 2013).

2. Methodology

The research follows a complex methodological and systematic approach to analyze how effective the detectors are for content generated by AI and what their limitations are. This analysis aims to provide a clear understanding of the main factors that lead to the emergence of errors and to highlight the shortcomings in how these systems interpret various literary, academic, religious texts, etc. The study tries to illustrate that these errors are not random but are

controlled by linguistic, structural, and contextual parameters. Thus, the applied methodology is based on 3 stages that we will see next:

2.1. Text selection and prepare

The corpus of texts examined in this research was selected with caution so as to maximize stylistic, temporal, and linguistic variation in poetry. The corpus includes:

– conventional classical poems, such as Mihai Eminescu's Luceafărul and Thomas Moore's The Last Rose of Summer poems widely recognized for literary and cultural significance, exemplified by elaborate expression and outdated vocabulary.

– translated texts, such as Tomorrow, at Dawn, the translation of Demain, dès l'aube by Victor Hugo, chosen to determine how translation affects detection by AI and whether changes in language affect categorization.

– modern and contemporary poetry, such as new authors, to compare today's style of writing with that of earlier times, and whether AI detectors are more sensitive to contemporary language than they are to classical forms.

The options were strategically selected to offer a foundation for comparative analysis across many styles, languages, and eras. One of the goals of the research is to attempt to find out whether there exists a correlation between the structural complexity of a poem and the likelihood of being caught as AI-generated.

2.2. Application of the AI Detector

Each poem within the corpus was analyzed individually using a well-known AI detection program called GPTZero, designed specifically to indicate whether or not content has been generated artificially. Each text was assigned a probability score by the system that indicated the level to which the content was being read as machine written. All results were recorded and compared, with particular attention paid to how the detector responded to linguistic and stylistic diversity.

To reduce external variables and ensure the analysis holds, each poem was tested multiple times with different settings within the AI program. This ensured that any inconsistency due to algorithmic variation was reduced to the barest minimum, and the results were consistent across trials.

2.3. Statistical and Stylometric Analysis

In an effort to find out how misclassification had taken place, a more detailed look at the information obtained was carried out. This included: Assessing the impact of translation on AI scoring, by determining whether

translated poems were more likely to be identified as artificially generated due to inherent structural alterations derived from translation.

Examining stylistic aspects of ancient and modern poems to determine which common characteristics would influence AI detection such as line length use of symbolism, or word repetition (Floridi, 2019).

Stylometric examination using natural language processing techniques with the intention of identifying whether certain linguistic patterns or configurations have measurable impacts on classification. This included frequency analysis, sentence complexity, and grammatical form distribution throughout the poems (Burrows, 2002).

Statistical modeling of results to uncover repeated patterns and possible bias in the AI system's evaluation of literature.

3. Results

The findings from the AI detector test shows drastic discrepancies in the categorization of both classical and modern literary pieces, indicating the present shortcomings of AI models in determining the authenticity of poetry. Classical literature, rich in rich grammatical structures, sophisticated metaphors, and old English words, was commonly confused with the machine-generated type in our research.

For instance, the Romanian poem Luceafărul by Mihai Eminescu (see Fig. 1) was labelled as AI-generated with a 96% probability, reflecting how the system is attempting to fail in the correct estimation of the depth of the classical poetic structure and the figure of speech of the author. The same outcome happened with the English romantic poem The Last Rose of Summer by Thomas Moore (see Fig. 2), with the result receiving a 100% probability of being AI-generated. This means that the AI system misreads the advanced stylistic markers and rhetorical flourishes typical of Romantic-era writing.

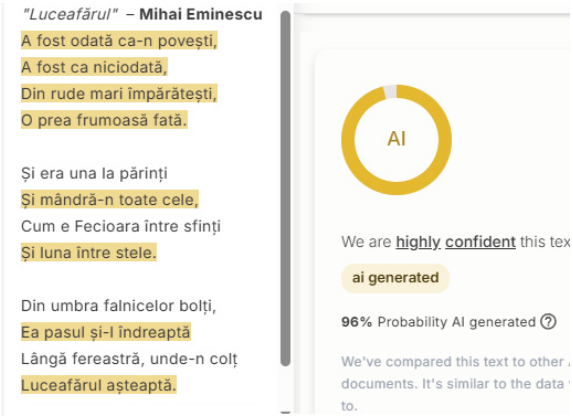


Fig. 1 – Poem Luceafărul.

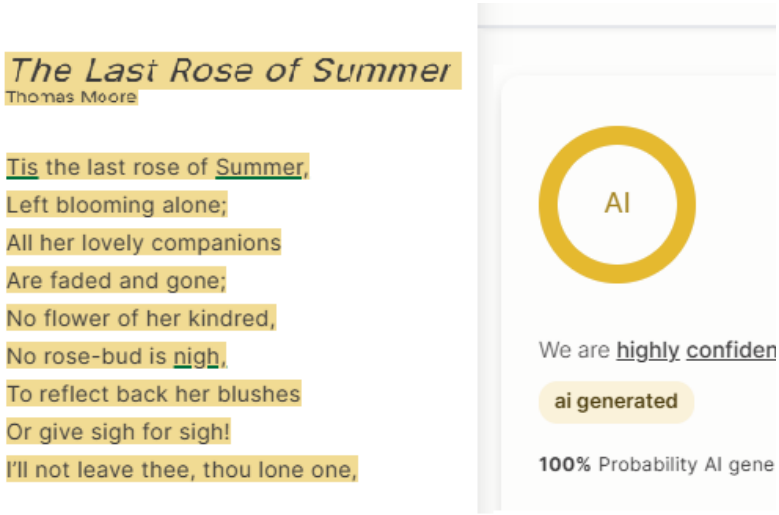


Fig. 2 – Poem The last rose of summer.

Literary translations also constituted a challenging category for AI classification. In our test, Victor Hugo's poem Demain, dès l'aube (Fig. 3) was labeled as AI-generated with a likelihood of 68% in the original French. However, its English translation–Tomorrow, at Dawn (see Fig. 4) was given the much higher score of 95%. This considerable disparity shows that the stylistic and structural changes effected in translation such as simplified syntax, altered poetic rhythm, or replacement of original vocabulary cause a significant difference in how AI systems recognize the authenticity of a text.

For contemporary texts, the results were mixed, with yet more challenges facing AI detectors in marking as such the actual literary work and what seems to be artificially generated language. In the poem "The Call of Jesus" by Arsenie Boca is a good example and was marked as AI-generated with 86% certainty. This misclassification may be due to the repetitive structures of the text very typical for religious texts and its heavy reliance on figurative language both tending to cause false positives.

These findings point to fundamental constraints in the manner in which AI systems are evaluating literary work, and consequences are severe in scholarly and editorial environments. When legitimate works are inaccurately labeled as machine-generated, it could cast their legitimacy into question, damage the reputation of an author, and introduce excessive doubt into scholarly examinations and publication discussions.

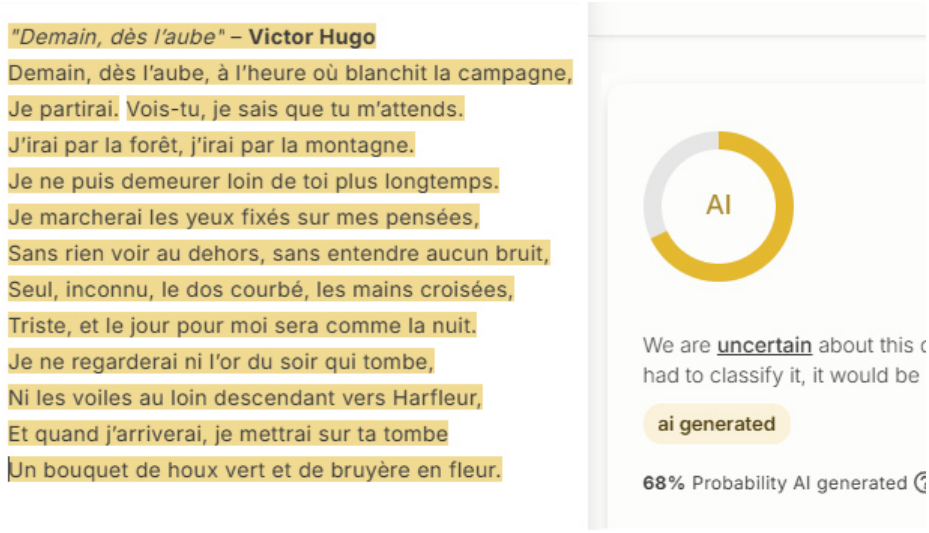


Fig. 3 – Poem *Demain, dès l'aube.*

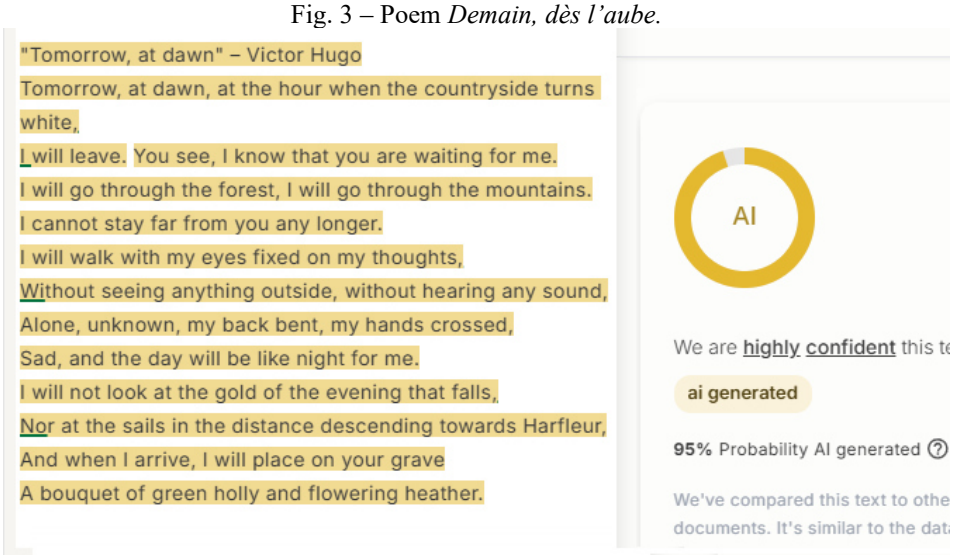


Fig. 4 – Poem *Tomorrow, at Dawn.*

All detection results of this study are presented in Table 1, where each text is assigned, a percentage indicating the likelihood of being AI-generated. In light of the reality that the majority of academic journals today already accommodate a highest of 20%, it thus becomes clear that of the 20 texts at hand, only 3 would make the cut under the accepted criteria while 17 would be rejected, despite never having been created through any form of artificial intelligence.

Table 1

Probability Percentage of Each Text Being AI-Generated

	Analyzed Text	AI Probability (%)	Observations
1	Tomorrow, at Dawn - Victor Hugo (EN)	95	The English translation receives a high AI probability score
2	Demain, dès l'aube - Victor Hugo (FR)	68	The original French version is less affected
3	Rhyme LIII - Gustavo Adolfo Bécquer (EN)	2	Authentic Spanish text detected as human
4	Rima LIII - Gustavo Adolfo Bécquer (ES)	26	The English version receives a lower AI probability score
5	Luceafărul - Mihai Eminescu (EN)	81	The English translation is detected as AI
6	Luceafărul - Mihai Eminescu (RO)	96	The original Romanian text has a very high AI probability
7	If - Rudyard Kipling (EN)	81	The English version detected as AI
8	Dacă - Rudyard Kipling (RO)	13	The Romanian version detected as human
9	Daffodils - William Wordsworth (EN)	100	Classic English poem incorrectly identified as AI
10	Narcise - William Wordsworth (RO)	80	The Romanian translation receives a high AI probability
11	Amurg - George Bacovia	86	Classic Romanian poem detected as AI
12	Agonie - Octavian Goga	82	Classic Romanian poem detected as AI
13	Amurgul Gândurilor - Emil Cioran	0	Philosophical essay correctly identified as human
14	Adio - Ana Blandiana	38	Contemporary poem detected with uncertainty
15	Himeră - Claudia Millian	37	Contemporary poem detected with uncertainty
16	Chemarea lui Iisus - Arsenie Boca	86	Religious text detected as AI
17	Sfânta Evanghelie după Matei - Bible	70	Bible detected as AI (70%)
18	Ultimul trandafir al verii - Thomas Moore (RO)	57	The Romanian translation of a classic poem detected as AI
19	The Last Rose of Summer-Thomas Moore (EN)	100	The original English poem detected as AI
20	One Too Many Mornings - Bob Dylan	79	Folk song detected as AI

4. Conclusions

Following this study, several aspects can be stated regarding the reason why the texts I analyzed were falsely reported, of course, as being generated by artificial intelligence. Therefore, a first premise would be the aspect that the texts use a formal and academic tone, employing an indirect expression, which leads to an impersonal tone. Furthermore, the words in the analyzed texts are chosen precisely, specifically, and clearly, avoiding colloquial language; of course, this is how it is, we are still talking about famous texts with massive academic significance and that is why artificial intelligence perceives it as inhuman, rather generated by an algorithm. It practically confirms that these analyzed works are truly literary masterpieces, created by minds with an exaggerated level of intelligence, comparable to an algorithm in today's world. But the correct question is: can the enlightened minds of today, who can think of texts in such a way, be accused of appropriating a text generated by artificial intelligence? A second premise is the fact that the analyzed texts have a correct grammatical structure, with transitional expressions that maintain a logical and coherent flow, aspects that can mislead the detectors of content generated by artificial intelligence.

As a result of this study, we can acknowledge that at this moment, the use of programs for detecting the use of artificial intelligence in generating academic text should be avoided. The results of such a program should, in no way, represent a decision-making factor regarding the acceptance or rejection of an academic paper by editors, nor in any other academic literary environment. A

false positive result on a literary text could have consequences for authors who have not actually used any AI tools. We can admit that these tools for detecting texts generated by AI could only be used as a guideline, and the final decision on the result should be made only after a proper verification by a human factor, in order to avoid false accusations.

REFERENCES

Andrews T.L., *Digital Philology and Manuscript Studies*, Literary and Linguistic Computing, vol. 34, nr. 2, pp. 173-192 (2019).

Brown T. et al., *Language Models are Few-Shot Learners*, arXiv preprint arXiv:2005.14165 (2020).

Burrows J., *Computational stylistics and authorship attribution*, Literary and Linguistic Computing, vol. 17, nr. 3, pp. 267-287 (2002).

Craig H., Kinney A., *Shakespeare, Computers, and the Mystery of Authorship*, Cambridge University Press, Cambridge (2009).

Eubanks V., *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York (2018).

Floridi L., *Artificial Intelligence and the 'Good Society*, Springer, Cham (2019).

Hoover D.L., *Stylometry, Text Analysis, and Literary Interpretation*, Literary and Linguistic Computing, vol. 20, no. 4, pp. 403-420 (2018).

Juola P., *Stylometric Analysis and Authorship Attribution*, Foundations and Trends in Information Retrieval, vol. 1, no. 3, pp. 233-334 (2008).

Kestemont M. et al., *Computational Authorship Verification Methods*, Journal of Digital Scholarship in the Humanities, vol. 34, no. 2, pp. 341-356 (2019).

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L., *The ethics of algorithms: Mapping the debate*, Big Data & Society, vol. 3, nr. 2 (2016).

Moretti F., *Distant Reading*, Verso Books, Londra, 2013.

Noble S.U., *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, New York (2018).

Organisciak P., Capitanu B., Underwood T., Downie J.S., *Access to billions of words: An overview of the HathiTrust Research Center Data and Tools*, Proceedings of the Joint Conference on Digital Libraries (2017).

Moore T., *The Last Rose of Summer*, in: *Poets.org – Academy of American Poets*. available at https://poets.org/poem/last-rose-summer, accessed on 19.04.2025.

Romanian poems available at www.poezii.romanianvoice.com/index.php, accessed on 22.04.2025.

LIMITĂRILE SOFTURILOR FOLOSITE ÎN DETECTAREA CONȚINUTULUI GENERAT DE CĂTRE IA

(Rezumat)

Această lucrare analizează discrepanțele dintre realitatea istorică a unor texte scrise cu zeci sau sute de ani în urmă și rezultatele oferite de programele moderne de detectare a conținutului generat de inteligența artificială (IA). Aceste discrepanțe sunt importante deoarece pot duce la erori în evaluarea autentică a textelor literare, afectând atât recunoașterea valorii culturale a acestor opere, cât și procesul editorial. Un astfel de fenomen poate contribui la o percepție eronată asupra autorilor clasici și la o lipsă de încredere în validitatea literaturii transmise de-a lungul secolelor. În mediul academic, unele texte clasice analizate cu astfel de detectoare IA au fost marcate eronat ca fiind generate artificial, ceea ce a determinat o reevaluare inutilă a autenticității acestora sau chiar respingerea lucrării academice. Printr-un studiu de caz asupra mai multor poezii clasice și contemporane, analizate cu un astfel de instrument, observăm cum detectoarele IA pot clasifica eronat texte autentice ca fiind generate de IA. Aceste erori ridică probleme privind utilizarea acestor instrumente în evaluarea originalității conținutului literar, mai ales în contextul academic și editorial.