Analyzing Large Language Models' pastiche ability: a case study on a 20th century Romanian author

Anca Dinu
Faculty of Foreign Languages and Literatures
University of Bucharest
Romania
anca.dinu@lls.unibuc.ro

Andra-Maria Florescu
Interdisciplinary School of Doctoral Studies
University of Bucharest
Romania
andra-maria.florescu@s.unibuc.ro

Liviu P. Dinu
Faculty of Mathematics and Computer Science
University of Bucharest
Romania
ldinua@fmi.unibuc.ro

Abstract

This study evaluated the ability of several Large Language Models (LLMs) to pastiche the literary style of the Romanian 20th century author Mateiu Caragiale, by continuing one of his novels left unfinished upon his death. We assembled a database of novels consisting of six texts by Mateiu Caragiale, including his unfinished one, six texts by Radu Albala, including a continuation of Mateiu's novel, and six LLM generated novels that try to pastiche it. We compared the LLM generated texts with the continuation by Radu Albala, using various methods. We automatically evaluated the pastiches by standard metrics such as ROUGE, BLEU, and METEOR. We performed stylometric analysis, clustering, and authorship attribution, and a manual analysis. Both computational and manual analysis of the pastiches indicated that LLMs are able to produce fairly qualitative pastiches, without matching the professional writer performance. The study also showed that ML techniques outperformed the more recent DL ones in both clusterization and authorship attribution tasks, probably because the dataset consists of only a few literary archaic texts in Romanian. In addition, linguistically informed features were shown to be competitive compared to automatically extracted features.

1 Introduction

The LLMs' capacity to imitate art is ever increasing in all creative domains. In literature, their ability to mimic the style of an author, of a character, of a literary genre, or of an epoch constitutes a vibrant research area with intriguing topics such as role-play (Wu et al., 2024), storytelling (Xie et al., 2023), creative writing (Chakrabarty et al., 2024). Since machine generation of literary pastiches of human authors raises ethical concerns due to the possibility of LLM-generated texts to pass as the work of human writers, Silva et al. (2024), research on the LLMs' ability to imitate a given author's style is much needed.

The term pastiche has a long history. It originates from the Italian *pasticcio*, meaning a mixture of meat and pasta turned into a pie. This food analogy suggests that the pastiche involves mixing available (recognizable) elements into a new thing, but without a new substance (Greene et al., 2012). Until the 20th century, the term had a negative connotation of a lack of creativity. Later, in theories of postmodernist literature, the term acquires its current meaning of an homage of past styles in the form of a deliberate imitation or blending of prior works of art, such as painting, architecture, design, sculpture, movie, music, poetry, or literature (Ayar, 2022). It consists of acknowledged borrowings of style, words, phrases, or motifs of previous authors, genres, or periods. The intention of pastiche is not mockery or forgery, but rather an open reference to the original (McArthur et al., 1996; Hutcheon, 2000), most often paying it a tribute. Some examples of literary pastiches are: extending a series when an author has died (like the Sherlock Holmes series, produced long after Sir Arthur Conan Doyle's death) or allowing fans to play with the narrative as in the case of fan fiction (like E L James' "Fifty Shades of Grey", the fanfic inspired by Stephenie Meyer's "Twilight").

In this paper, we investigate the LLM's capacity to pastiche an author style. To do so, we propose a case study on an intriguing literary pastiche case from Romanian 20th century literature. This choice was motivated by the existence of a pastiche novel authored by a professional writer who tried to imitate the style of another author, which naturally constitutes a golden standard for comparison of the machine-generated pastiches.

The original novel was written by Mateiu Caragiale (1885–1936), a Romanian Symbolist and Decadent writer recognized for his role in modernizing the Romanian literary language, through his unique voice, stylistic innovation, lexical baroque richness, elaborate syntax, poetic language, and

Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities, pages 20–32
May 3-4, 2025

focus on mood over plot. In his last seven years of life, he authored the novel *Sub pecetea tainei* (Under the seal of secrecy) without finishing it. Some decades later, in the 1970s, the rumor that the continuation of the novel was found spread in Romanian literary circles, passing for a short time as a possibly genuine ending of Mateiu's last novel, due to its very similar writing style. The debate was settled by Radu Albala, the actual author of the continuation entitled *În deal, pe Militari* (On the Militari hill). He revealed that his goal was precisely to continue the original novel in a style so similar to the original writer, as to pass as Mateiu's text for human experts. Radu Albala (1924-1994) was one of the closest stylistic followers of Mateiu Caragiale, among others like Eugen Bălan and Alexandru George, who also wrote continuations of the unfinished novel of Mateiu, as a stylistic exercise (Dinu et al., 2012).

The rest of the paper is structured as follows: the next section presents related work; the next one describes the data in detail. The Analysis section is divided in two subsections: one for the computational analysis, comprising evaluation metrics between the original and the pastiches, stylometric analysis, and automatic methods such as pastiche clustering and authorship predictions, and the other focusing on human interpretation. We summarize the findings of our study in the Conclusions section.

2 Related Work

A thorough survey of stylometry or authorial style, comprising techniques, tools, and algorithms can be found in (Neal et al., 2017a).

The methodology of stylometry centered on authorship or style debates of old texts is well established and used in numerous recent research, like (Kawasaki, 2022), who performed stylometric analysis based on POS and n-grams on Amadís de Gaula and its sequel Sergas de Esplandián medieval Spanish chivalric romances, or (Kawasaki, 2023) who focuses on authorship attribution with POS and n-grams stylistic features on 15th century *Tirant lo Blanc*, or (Miyagawa et al., 2024), who analyses the (word embeddings) semantic similarity and intertextuality of the Vedic Sanskrit corpus.

In the field of the more recent LLM-generated texts, a comprehensive literature review of authorship attribution Huang et al. (2025) categorizes four representative problems: human-written text attribution, LLM-generated text detection, LLM-generated text attribution, and human-LLM coauthored text attribution.

LLMs can be prompted to generate any kind of creative text, in any manner. For instance, Silva et al. (2024) prompted ChatGPT to forge a novel and not the author's style. Another example is the prolific domain of creative writing. To give a very recent instance, Chakrabarty et al. (2024) evaluated the creative writing abilities of three LLMs and ten humans, instructing them to create a story based on a prompt that included the summary of a novel. The results showed that the LLMs performed worse than humans. Also, they used LLMs to asses the quality of the generated tests, but their evaluations correlated poorly with human judgment. Kumarage and Liu (2023) and Muñoz-Ortiz et al. (2024) compared LLMs and humans writing style on news articles, finding that there are relevant distinctive features between the two. Durward and Thomson (2024) investigate vocabulary usage for AI and human-generated text in news articles and creative writing, noting thematic differences between them. Reinhart et al. (2024) identified systematic differences between LLMs and humans on different register texts. Chen and Moscholios (2024) explored LLMs capacities of imitating a person's language style. Bhandarkar et al. (2024) proposed the task of emulating human style with LLMs on blog posts.

Previous work on Romanian 20th century writers Mateiu Caragiale and Radu Albala (Dinu et al., 2008) focused on authorship identification for Albala's pastiche of Mateiu's unfinished novel, using stop words rankings. Another similar research (Dinu et al., 2012) measured the style similarities between Mateiu's writing and the writing of his followers, who tried to mimic or pastiche him (Albala, Agopian, Bălan, and Iovan), finding that they are closer in style to each other than to Mateiu.

3 Data

We obtained the six original novels by Mateiu Caragiale, published as volume chapter of the book "Craii de curtea veche", from WikiSource. For Radu Albala, we obtained the six novels from a Publishing House, for research purposes.

The pastiches generated by the LLMs were obtained by few-shot prompting, providing them with the last unfinished novel written by Mateiu Caragiale, *Sub pecetea tainei*. We used the following prompt to ask the LLMs to generate a pastiche that continues it: *You are Mateiu Caragiale, a Ro-*

manian writer, son of I.L. Caragiale. Continue the plot with 18000 characters from the short story Sub pecetea tainei! Here is an example of how Mateiu wrote: "...". The choice of the generated text length is motivated by the intention to match the length of 18528 characters of Albala's pastiche *In deal pe Militari* that continued Mateiu's *Sub pecetea tainei*, so as to directly compare the LLMs generated texts with the professional writer's pastiche.

We used six publicly available LLMs for this pastiche generation task: ChatGPT4o[1], Claude Haiku[2], Gemini 1.5 pro[3], Qwen 2.5 72b instruct[4], Wizzard LM2 8x22b, and Llama 3.1 70b Turbo (both accessed via Deepinfra chat platform[5]). For Gemini, we deactivated all safety settings, as this feature was available and since negative sentiments have been shown to correlate with artistic creativity (Akinola and Mendes, 2008). We did not change any other parameters of the models, like top-p or temperature, as we focused on their default generative capacities.

We manually inspected the texts and cleaned them accordingly. We removed any special characters. We standardized the dialogue marker, since in some texts a small dash was used and some of the LLMs used the English standard quotation marks, replacing them all by the standard Romanian Em-dash. We also cleaned any page number, footnote mention, or others.

The data set is well balanced, in terms of the number of examples per author and of the text length. We give the name of all the novels by human authors and the data statistics in table 1.

4 Analysis

4.1 Computational approach

In this section, we will employ a set of computational methods to analyze the pastiche dataset: quantitative analysis that includes evaluation metrics between the original and the pastiches, stylometric analysis, and automatic methods such as pastiche clustering and authorship predictions.

4.1.1 Experimental setup

All automated experiments employed zero- or few-shot prompt engineering with coding assistance from Claude haiku and ChatGPT4. This was a trial-and-error process until we received the desired results. We experimented with both traditional Machine Learning (ML) techniques and more advanced Deep Learning (DL) approaches like transformers. The experiments were performed with Python in Google Colab using libraries like: spaCy, transformers, nltk, sklearn, numpy, pandas, matplotlib.

4.1.2 Automatically evaluating pastiche generation by standard metrics

The most straightforward way to compare two documents is to use standard assessment measures such as: ROUGE, BLEU, and METEOR, which are language independent. We computed these metrics for the original novel by Mateiu *Sub pecetea Tainei*, as the reference text, and all six LLM generated texts that were supposed to pastiche it, plus Albala's *În deal pe Militari* that continued Mateiu's novel. In addition, we calculated two other measures, Diversity and Perplexity, to assess the quality of the generated texts. For comprehensive surveys on the use of automated metrics for Natural Language Generation see (Celikyilmaz et al., 2021) and (Schmidtova et al., 2024).

ROUGE score (Lin, 2004) measures the overlap between n-grams of the reference text and the generated text. The higher the value, the more the two texts overlap, so they are more similar in terms of structural alignment. However, ROUGE does not account for words with similar meaning, as it does not mind semantics, and it sticks solely to n-grams containing identical words. Moreover, this evaluation metric focuses only on recall, that is, on how much the words/n-grams in the reference text appear in the model generated text. Complementary, the BLEU score (Papineni et al., 2002) focuses on precision: how much the words/n-grams in the model generated text appear in the reference.

METEOR (Banerjee and Lavie, 2005) is a metric specifically designed to address the shortcomings of ROUGE and BLEU. Firstly, it computes the score as the harmonic mean of the n-gram precision and recall, assigning a higher weight to recall than to precision. Secondly, METEOR considers morphological variations of words and synonyms, thus measuring also semantic similarity.

ROUGE, BLEU, and METEOR were originally designed to score the similarity between an original human text and a machine-generated one, for specific tasks such as automatic translation, summarization, or rephrasing. Nevertheless, they have

[1] https://chatgpt.com/
[2] https://claude.ai/chat
[3] https://aistudio.google.com/prompts/new_chat
[4] https://huggingface.co/spaces/Qwen/Qwen2.5
[5] https://deepinfra.com/chat

Author	Title	Length (characters)
Mateiu Caragiale	Întâmpinarea crailor	32,137
	Cele trei hagialâcuri	48,924
	Spovedanii	58,132
	Asfințitul Crailor	67,388
	Remember	37,248
	Sub pecetea tainei	63,223
Radu Albala	Propylaën Kunstgeschichte	21,514
	La Paleologu	89,100
	Niște cireșe	17,803
	Sclava iubirii	42,769
	Femeia de la miezul nopții	112,558
	În deal, pe Militari	18,528
LLMs (Sub pecetea tainei)	ChatGPT4o	18,855
	Claude Haiku	17,702
	Gemini 1.5 pro	17,011
	Llama 3.1 70b Turbo	18,574
	Qwen 2.5 72b instruct	17,845
	Wizzard LM2 8x22b	17,510

Table 1: The dataset

also been used subsequently for evaluating general purpose automatic text generation. Although initial research reported that they correlate well with human judgments (Agarwal and Lavie, 2008), more recent work (Caccia et al., 2020) pointed out that texts with very high scores, while perfectly grammatical, can lack semantic or global coherence and can present a poor narrative flow.

To assess the quality of the generated texts, without comparison with the reference text, we employed Diversity and Perplexity measures, which quantify the variety, and the naturalness of the language, respectively. Diversity measures the lexical richness of the generated text by calculating the ratio of unique n-grams to the total n-grams. Higher diversity implies the generation of more varied and creative content. Perplexity measures the uncertainty of the language model in predicting the next word, thus, lower perplexity indicates better fluency and less uncertainty in text generation.

We first lemmatized the Romanian texts with SpaCy, preserving stop words and punctuation, and converting it all to lowercase, then we used chunking to dynamically handle long text. To compute ROUGE and BLEU scores, we used nltk libraries. For METEOR we employed readerbench/RoBERT-base from HuggingFace to compute similarity between words and map them if they cross a certain threshold (set to an optimum 0.65), despite them not being the exact same word. The final METEOR score is a weighted F1 score, giving 9:1 weightage for precision over recall. To compute the Diversity metric we used bi-grams. Perplexity was calculated with the same pre-trained model and normalized to 0-1 interval values. The scores for all metrics are given in table 6 from the Appendix.

As illustrated in figure 1, the professional writer, Radu Albala, outperformed the six LLMs in mimicking the reference text. Albala obtained the highest ROUGE, BLEU, METEOR, and Diversity compared to the LLMs, meaning that his pastiche was the most fluent, the most similar to the original text, both grammatically and semantically, and had the richest vocabulary. Nevertheless, his absolute scores show that, while he successfully mimicked the writing style of Mateiu, his personal, original, writing style is still present.

In terms of Perplexity, Qwen obtained the lowest score, meaning a more predictable, natural writing style. However, there is a fine line between writing naturally and writing predictably and METEOR score cannot differentiate between the two. A writer is expected to write with naturalness, but not to have a very predictable wording.

The results reveal notable differences in the performance of the models across various evaluation metrics. ChatGPT achieves the best performance among the six LLMs, leading in ROUGE, BLEU,

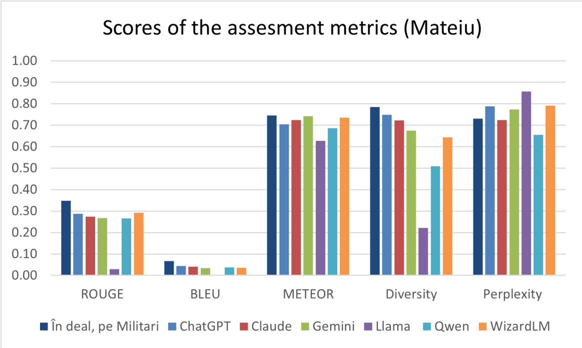

Figure 1: Assessment measures for the similarity of Mateiu's "Sub pecetea tainei" with its pastiches.

METEOR, and Diversity scores. Claude, Gemini, and WizardLM also perform competitively. Qwen has the lowest perplexity, indicating that it might generate the most predictable wording, with a moderate diversity. Llama was the lowest performing model, indicating heavy repetition, lack of vocabulary richness, poor fluency and unnatural phrasing.

4.1.3 Stylometric analysis

To further analyze style similarities between the considered texts, we performed quantitative analysis of literary style, or stylometry, based on linguistic features such as word frequency, sentence length, or syntactic patterns (Neal et al., 2017b). We used Linguistic Inquiry and Word Count (LIWC-22) (Boyd et al., 2022) and Python scripts. LIWC is a text analysis tool based on socio-linguistic features, psychologically motivated, which uncovers emotional, cognitive, and structural components. We extracted its default 86 features available for the Romanian dictionary (Dudău and Sava, 2020; Crudu, 2024) from all 18 texts in our dataset. We manually trimmed the feature set to fit our specific purposes (authorship-centered), ending up with only 34 relevant ones, structured into 3 groups: part of speech frequencies (functional words included), punctuation, and sentiments, shown in tables 7, 8, and 9 from the Appendix, respectively.

We next experimented with traditional ML methods to see whether the three text categories, Albala, Mateiu, and LLMs can be automatically clustered together, considering only the 18 vectors containing the linguistically informed selected features. We used the agglomerative clustering algorithm, and Principal Component Analysis (PCA) to reduce the space to 2 dimensions, for convenient visualization. It turns out that the selected features extracted with LIWC-22 were informative enough to cluster together the three categories, as shown in figure 2. Moreover, the pastiche *În deal, pe Militari* is the closest of all Albalas's texts to Mateiu's cluster (centroid) and the farthest from its own class.

Since the clusterization results suggest that the texts might be grouped together automatically by their authors, one legitimate question is if one can automatically predict the authorship of the pastiche correctly and with what probability. To test that, we trained a Support Vector Machine classifier (SVM) on five texts written by Albala, five written by Mateiu, and five pastiches generated by LLMs, in two scecnarios: two classes prediction (Mateiu, and Albala) and three classes prediction (Mateiu, Albala, and LLMs). We fed the model the original novel written by Mateiu and the pastiche written by Albala and asked it to predict to what class each text belongs to, and give the associated probabilities. The results for three classes prediction are shown in table 2. One can see that both novels were correctly predicted to have been written by their actual authors: *Sub pecetea tainei* to Mateiu, with 52.10 % probability, and Albala's pastiche to himself, with 57.15 % probability. When we dropped the LLM class, the prediction performance increased, as illustrated in table 3: *Sub pecetea tainei* was attributed to Mateiu with 73.96 %, and *În deal, pe Militari* to Albala, with 72 %. These results surpass the previous predictions in (Dinu et al., 2008), where the authors reported that a SVM model with linear kernel correctly attributed the original to Mateiu with a probability of 62.56 %, and the pastiche to Albala with a probability of 50.56 %.

We also computed with LIWC the language style matching (LSM) that measures the degree of writing style matching by calculating similarity in the use of function words. While the LSM score between Albala's pastiche and Mateiu's original novel is 0.66, the LSM scores between LLM generated pastiches and the original novel range between 0.47 and 0.63. This shows once again that the professional writer managed to get closer to Mateiu's writing style than the LLMs.

4.1.4 Clusterization and authorship attribution

While in section 4.1.3 we automatically clustered and predicted the authors of the pastiches based only on vectors of extracted linguistically informed features, in this section we automatically cluster

Sub pecetea tainei	Authorship probabilities (based on LIWC features)
Mateiu	**52.10** %
Albala	40.28 %
LLMs	7.62 %
În deal, pe Militari	**Authorship probabilities (based on LIWC features)**
Mateiu	28.09 %
Albala	**57.15** %
LLMs	14.76 %

Table 2: three classes authorship prediction for original and pastiche texts, based on LIWC features

Sub pecetea tainei	Authorship probabilities (based on LIWC features)
Mateiu	**73.96** %
Albala	26.04 %
În deal, pe Militari	**Authorship probabilities (based on LIWC features)**
Mateiu	28 %
Albala	**72** %

Table 3: two classes authorship prediction for original and pastiche texts, based on LIWC features

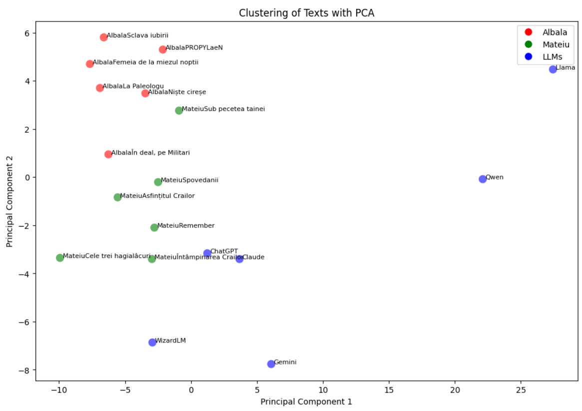

Figure 2: Clusterization based on LIWC features.

and predict the authors of the pastiches employing ML and DL approaches that use the entire texts as input. We kept all punctuation and stop words, since in authorship studies they have been proven to best distinguish between different authors (Dinu et al., 2008, 2012).

To cluster the 18 files into the 3 groups (authored by Mateiu, Albala, or LLMs), we used k-means and agglomerative clustering algorithms, both employing the Euclidean distance. We only give here the results obtained with agglomerative clustering, which were more clear-cut than the ones obtained with k-means, probably because it does not assume a spherical shape of the clusters, like k-means does. We experimented with three ways of extracting the features from the texts: Term Frequency-Inverse Document Frequency (tf-idf), BERT-embeddings Romanian version (Dumitrescu et al., 2020), and hybrid (tf-idf plus Romanian BERT). The performance of the clusterization based on the Romanian BERT embeddings was the poorest, most probably because of the archaic Romanian used in the text, unseen by the model in the training data. Moreover, the hybrid approach gave the same results as the tf-idf one. Consequently, we only report here the results based on tf-idf method.

The graphical representations of the clusters were obtained using PCA to initially reduce the dimensionality of the data, followed by Uniform Manifold Approximation and Projection for Dimention Reduction (UMAP) [6] to refine the initial PCA and provide a clearer 2D visualization.

For the tf-idf vectorization approach, we used spaCY for Romanian to preprocess the data, including lemmatizing it. The resulting tf-idf vectors were scaled using StandardScaler to standardize the data before clustering. Figure 3 shows two clusters representation (Mateiu, and Albala), and figure 4 displays the three-clusters representation (Mateiu,

[6] https://umap-learn.readthedocs.io/en/latest/

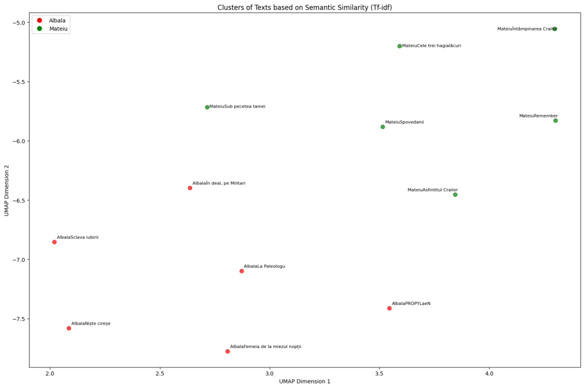

Figure 3: Tf-idf clustering of Mateiu and Albala texts

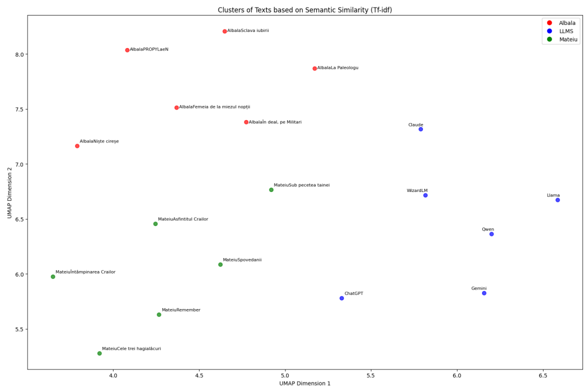

Figure 4: Tf-idf clustering of Mateiu, Albala, LLM texts

Albala, and LLMs). In both of them, the pastiche *În deal, pe Militari*, while correctly represented in it's own Albala cluster, is the closest to the original Mateiu's novel that it supposed to pastiche. The three group clustering shows that LLM generated texts are a clearly defined group, away from human written texts.

As in the case of using the linguistically informed features extracted with LIWC, we experimented next with two classes and three classes authorship prediction. All models used chunking with overlap, meaning that the text is split into overlapping chunks of 100 words with 50-word overlap, ensuring contextual continuity across fragments. The final predictions are based on the average probabilities of all chunks. In total, we experimented with five models: SVM with tf-idf vectorization, predictions based on Rank Distance (Popescu and Dinu, 2008), Romanian BERT (Dumitrescu et al., 2020), readerbench/RoBERT-base (Masala et al., 2020), and the sentence transformer MiniLM-L12-v2 (Reimers and Gurevych, 2019). After we trained the models, we asked them to predict the unseen original and pastiche texts.

To validate the SVM tf-idf model, we experimented with several hyperparameters by using Grid Search Cross-Validation, with 5 fold cross-validation for robustness. The most competitive model had a linear kernel, limiting the number of features to 5000 for efficiency. Its train/test split ratio was the 80/20.

For the model based on Rank Distance, we used the comprehensive stop words list for Romanian language [7]. We trained a SVM classifier with linear kernel to distinguish between different authors on feature vectors extracted from the texts, based on the frequency distribution of both stop words and content words. The hyperparameters were tuned using Grid Search Cross-Validation, in the same manner as for the SVM tf-idf model.

We further experiment with three transformers: two variants of Romanian BERT (Dumitrescu et al., 2020; Masala et al., 2020), at word level, and a multilingual sentence transformer, MiniLM-L12-v2 (Reimers and Gurevych, 2019). Since the prediction performance of the SVM models trained on embeddings obtained with the three transformers was poor, we changed the ML technique to Logistic Regression in all three cases.

The best overall performance was achieved by the Rank Distance with function words, which always assigned the right classes, with the highest probability in three out of four cases, as shown in tables 4 and 5. It has been slightly surpassed by the SVM tf-idf model only in the case of the three class prediction of the original novel. The second best model was SVM tf-idf. All transformer-based models underperformed, in comparison with the traditional ML methods. Moreover, the two Romanian BERT models misclassified the pastiche as being written by Mateiu. The multilingual sentence transformer was the only one to correctly classify all the cases, but with much lower probabilities than the ML approaches.

As in the case of clusterization, the better performance of ML methods over the transformers might be explained by the fact that the Romanian language used in the literary texts was non-standard and archaic, a type of language not seen in the training data for the transformers.

The multilingual sentence transformer's higher performance compared to the Romanian word-embeddings BERT models could result from the

[7]https://github.com/stopwords-iso/stopwords-ro?tab=MIT-1-ov-file

training data of the sentence transformer, which is more varied in terms of language versions, leading to better generalization ability. Also, BERT transformers need larger datasets to generalize well, while sentence transformers are better fitted to train on small datasets. Lastly, sentence transformers capture the meaning of entire sentences, making them ideal for text prediction where chunk-based embeddings are averaged.

4.2 Qualitative analysis

In Digital Humanities (DH), where datasets are often sparse, nonstandard, and/or in a low resourced language, the computer-assisted approach is the most appropriate. This means that computational methods provide valuable insight to the humanist from the data at hand, but the final inspection and interpretation should be human. Since we deal with Romanian literary texts with 20th century vocabulary and structure, a manual analysis of the human and LLM generated pastiches was in line. In doing this, we focused on the following criteria: linguistic and technical quality (grammar, coherence, narrative structure), stylistic similarity with the original (similar vocabulary, figurative language use, mood), and original contributions.

All LLM generated texts contain grammatical errors in various degrees. The least grammatical errors were made by ChatGPT, while the most errors were made by Qwen. Some systematic mistakes that occurred frequently were: feminine gender disagreement, missing or erroneous diacritics, spelling errors, and various morpho-syntactic errors, most notably related to declension, conjugation, reflexive pronouns, and accusative case assignment.

Most LLMs are coherent and easy to follow, except for Llama, which is very repetitive in terms of entire paragraphs and sentence beginnings with present perfect tense, first-person singular. Qwen is also fixated on repeatedly using this tense.

In the original novel, Mateiu changed back and forth the narrative perspective between two characters, with first-person point of view. The human pastiche maintains this feature, while LLMs were largely confused by it and couldn't successfully imitate this. WizardLM used only third-person point of view, while Qwen, Gemini, and Llama use only one first-person narrator. ChatGPT and Claude managed to keep a dual narrative perspective, but wrongfully switched between the two.

The vocabulary used by the LLMs was well adapted to the time of the narrative, but the word forms used were the standard contemporary Romanian ones, in contrast to the original novel, where a considerable amount of words appear in their archaic form. Moreover, the original text abounds in foreign language quotes and expressions, mostly in French, Latin, and German. The LLMs generally failed to include expressions in languages others than Romanian, with the exception of Gemini, which inserted some French expressions, and of Claude, which used one Latin phrase.

Mateiu's original novel uses rich figurative speech (complex metaphors, epithets, comparisons, etc.), which ChatGPT, Gemini, and Claude successfully imitated. WizardLM overdoes it, its figurative speech seeming somehow forced. Qwen's figurative speech is rather simplistic, resembling middle school level homework, while Llama's seems closer to elementary school level.

While the original novel creates a mysterious detective fiction atmosphere, largely maintained in Albala's pastiche, all LLMs expressed their own nuances on the mood they created. ChatGPT expanded the original mysterious atmosphere towards mysticism; Claude brought a touch of positivism and symbolism; Gemini's pastiche presented thriller and realistic traits; Llama's pastiche seemed a hallucination; Qwen was the most faithful to the detective atmosphere of the original; finally, WizardLM created a mostly romantic atmosphere.

Most LLM generated texts had a happy ending. This might be explained by the LLMs' active filters. The only exception was Gemini, for which we turned off the filters, and which generated a story where the main feminine character died.

These observations correlate with similarity metrics scores, stylometric analysis, clusterization, and prediction, complementing each other's insights.

5 Conclusions

In general, LLMs generated fairly good pastiches, although without matching the quality of the human written pastiche. This is supported by all scores and methods used: similarity scores, stylometry, language style matching scores, clusterization, prediction, and manual inspection. Overall, traditional ML methods outperformed more recent DL ones. This happened because our data consisted in a few literary archaic text in Romanian, this kind of dataset being typical of DH. Nevertheless, a study focused on contemporary English could show bet-

Sub pecetea tainei	SVM (TF idf)	Rank distance (stop + content words)	BERT (Dumitrescu)	BERT (RoBERT)	Sentence transformer
Mateiu	70.98 %	**71** %	66.33 %	68.68 %	59.18 %
Albala	29.02 %	29 %	33.67 %	31.32 %	40.82 %
În deal, pe Militari	**SVM (TF idf)**	**Rank distance (stop + content words)**	**BERT (Dumitrescu)**	**BERT) (RoBERT)**	**Sentence transformer**
Mateiu	28.53 %	23.77 %	58.55 %	55.57 %	46.12 %
Albala	71.47 %	**76.23** %	41.45 %	44.43 %	53.88 %

Table 4: two classes authorship prediction for *Sub pecetea tainei* and *În deal, pe Militari*

Sub pecetea tainei	SVM (TF idf)	Rank distance (stop + content words)	BERT (Dumitrescu)	BERT (RoBERT)	Sentence transformer
Mateiu	**71.89** %	70.75 %	64.89 %	66.57 %	54.84 %
Albala	26.08 %	27.68 %	34.65 %	33.21 %	37.61 %
LLMs	2.03 %	1.57 %	0.46 %	0.21 %	7.55 %
În deal, pe Militari	**SVM (TF idf)**	**Rank distance (stop + content words)**	**BERT (Dumitrescu)**	**BERT) (RoBERT)**	**Sentence transformer**
Mateiu	31.15 %	20.27 %	55.82 %	57.14 %	39.77 %
Albala	67.04 %	**77.77** %	42.24 %	40.66 %	51.46 %
LLMs	1.81 %	1.96 %	1.94 %	2.19 %	8.78 %

Table 5: three classes authorship prediction for *Sub pecetea tainei* and *În deal, pe Militari*

ter performance of LLMs and of DL methods.

Finally, linguistically informed features proved to be competitive compared to automatically extracted features. Also, task-specific methods like Rank Distance similarity, known to perform well on authorship identification, outperformed general-purpose models.

Limitations

We only included in this study one of the writers who imitated Mateiu's writing style. In future work, we will expand the analysis to other Romanian authors considered followers of Mateiu Caragiale, like Ion Iovan, who created a diary fiction impersonating Mateiu, and others.

We also plan to increase the number of LLMs we used. Another research venue will be to experiment with different LLM parameters such as temperature, or top p, to investigate how the pastiche performance of LLM varies with these settings. Moreover, we are interested in further investigating the influence of prompt styles (like zero-shot, Chain-of-thought, Tree-of Thoughts, Retrieval-Augmented Generation) on the pastiche generation task, since in this study, we only used few-shot prompt type. Fine-tuning LLMs specifically for pastiche generation is another valuable research option to explore.

We consider other literary aspects worthy of further analysis, such as narrative pacing, character portrayal, Named Entities consistency (places, time, characters, etc.), references similarity, etc.

Ethics Statement

This research adheres to ethical standards regarding the use of literary works. Mateius's novels were written in the early 20th century, which makes them open source according to the Romanian copyright law (Law No. 8/1996 on Copyright and Related Rights), which grants protection for 70 years after the author's death. Albala's novels were obtained from a publishing house, ensuring that its use complies with legal and ethical guidelines. All excerpts used are for scholarly purposes, and proper attribution is maintained to respect intellectual property rights, following the provisions set forth in Law No. 8/1996 regarding fair use for educational and research purposes.

Moreover, we are not releasing the datasets to the public to prevent any unethical usage of the original and of LLM generated novels.

We respected all licensing agreements for all the software, libraries, and models we used.

References

Abhaya Agarwal and Alon Lavie. 2008. Meteor, m-bleu and m-ter: Evaluation metrics for high-correlation with human rankings of machine translation output. In *WMT@ACL*.

Modupe Akinola and Wendy Berry Mendes. 2008. The dark side of creativity: biological vulnerability and negative emotions lead to greater artistic creativity. *Personality & social psychology bulletin*, 34(12):1677–1686.

MUzzafer Zafer Ayar. 2022. How to cope with postmodern texts: Textual analysis of intertextuality, parody, and pastiche in reading postmodern texts. *JOURNAL OF MODERNISM AND POSTMODERNISM STUDIES (JOMOPS)*, 3(1):183–191.

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.

Avanti Bhandarkar, Ronald Wilson, Anushka Swarup, and Damon Woodard. 2024. Emulating author style: A feasibility study of prompt-enabled text stylization with off-the-shelf LLMs. In *Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024)*, pages 76–82, St. Julians, Malta. Association for Computational Linguistics.

Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, 10.

Massimo Caccia, Lucas Caccia, William Fedus, Hugo Larochelle, Joelle Pineau, and Laurent Charlin. 2020. Language gans falling short. In *International Conference on Learning Representations*.

Asli Celikyilmaz, Elizabeth Clark, and Jianfeng Gao. 2021. Evaluation of text generation: A survey.

Tuhin Chakrabarty, Philippe Laban, Divyansh Agarwal, Smaranda Muresan, and Chien-Sheng Wu. 2024. Art or artifice? large language models and the false promise of creativity. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery.

Ziyang Chen and Stylios Moscholios. 2024. Using prompts to guide large language models in imitating a real person's language style.

Mălina Crudu. 2024. Automatic detection of verbal deception in romanian with artificial intelligence methods. *Studia Universitatis Babeș-Bolyai Informatica*, 69(1):70–86.

Liviu Dinu, Marius Popescu, and Anca Dinu. 2008. Authorship identification of Romanian texts with controversial paternity. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Liviu P. Dinu, Vlad Niculae, and Maria-Octavia Sulea. 2012. Pastiche detection based on stopword rankings. exposing impersonators of a Romanian writer. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, pages 72–77, Avignon, France. Association for Computational Linguistics.

D. P. Dudău and F. A. Sava. 2020. The development and validation of the romanian version of linguistic inquiry and word count 2015 (ro-liwc2015). *Current Psychology*, 41:3597–3614.

Stefan Dumitrescu, Andrei-Marius Avram, and Sampo Pyysalo. 2020. The birth of Romanian BERT. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4324–4328, Online. Association for Computational Linguistics.

Matthew Durward and Christopher Thomson. 2024. Evaluating vocabulary usage in LLMs. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 266–282, Mexico City, Mexico. Association for Computational Linguistics.

R. Greene, S. Cushman, C. Cavanagh, J. Ramazani, and P. Rouzer. 2012. *The Princeton Encyclopedia of Poetry and Poetics: Fourth Edition*. Princeton Reference. Princeton University Press.

Baixiang Huang, Canyu Chen, and Kai Shu. 2025. Authorship attribution in the era of llms: Problems, methodologies, and challenges.

L. Hutcheon. 2000. *A Theory of Parody: The Teachings of Twentieth-Century Art Forms*. University of Illinois Press.

Yoshifumi Kawasaki. 2022. A stylometric analysis of amadís de gaula and sergas de esplandián. In *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*, pages 1–7, Taipei, Taiwan. Association for Computational Linguistics.

Yoshifumi Kawasaki. 2023. Revisiting authorship attribution of tirant lo blanc using parts of speech n-grams. In *Proceedings of the Joint 3rd International Conference on Natural Language Processing for Digital Humanities and 8th International Workshop on Computational Linguistics for Uralic Languages*, pages 16–26, Tokyo, Japan. Association for Computational Linguistics.

Tharindu Kumarage and Huan Liu. 2023. Neural authorship attribution: Stylometric analysis on large language models.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Mihai Masala, Stefan Ruseti, and Mihai Dascalu. 2020. Robert–a romanian bert model. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6626–6637.

T. McArthur, T.B. McArthur, and R. McArthur. 1996. *The Oxford Companion to the English Language*. Oxford Companions Series. Oxford University Press.

So Miyagawa, Yuki Kyogoku, Yuzuki Tsukagoshi, and Kyoko Amano. 2024. Exploring similarity measures and intertextuality in Vedic Sanskrit literature. In *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, pages 123–131, Miami, USA. Association for Computational Linguistics.

A. Muñoz-Ortiz, C. Gómez-Rodríguez, and D. Vilares. 2024. Contrasting linguistic patterns in human and LLM-generated news text. *Artificial Intelligence Review*, 57(10):265.

Tempestt Neal, Kalaivani Sundararajan, Aneez Fatima, Yiming Yan, Yingfei Xiang, and Damon Woodard. 2017a. Surveying stylometry techniques and applications. *ACM Comput. Surv.*, 50(6).

Tempestt Neal, Kalaivani Sundararajan, Aneez Fatima, Yiming Yan, Yingfei Xiang, and Damon Woodard. 2017b. Surveying stylometry techniques and applications. *ACM Comput. Surv.*, 50(6).

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Marius Popescu and Liviu Dinu. 2008. Rank distance as a stylistic similarity. volume 1, pages 91–94.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Alex Reinhart, David West Brown, Ben Markey, Michael Laudenbach, Kachatad Pantusen, Ronald Yurko, and Gordon Weinberg. 2024. Do llms write like humans? variation in grammatical and rhetorical styles.

Patricia Schmidtova, Saad Mahamood, Simone Balloccu, Ondrej Dusek, Albert Gatt, Dimitra Gkatzia, David M. Howcroft, Ondrej Platek, and Adarsa Sivaprasad. 2024. Automatic metrics in natural language generation: A survey of current evaluation practices. In *Proceedings of the 17th International Natural Language Generation Conference*, pages 557–583, Tokyo, Japan. Association for Computational Linguistics.

Kanishka Silva, Ingo Frommholz, Burcu Can, Fred Blain, Raheem Sarwar, and Laura Ugolini. 2024. Forged-gan-bert: Authorship attribution for llm-generated forged novels. In *Conference of the European Chapter of the Association for Computational Linguistics*.

Weiqi Wu, Hongqiu Wu, Lai Jiang, Xingyuan Liu, Hai Zhao, and Min Zhang. 2024. From role-play to drama-interaction: An LLM solution. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3271–3290, Bangkok, Thailand. Association for Computational Linguistics.

Zhuohan Xie, Trevor Cohn, and Jey Han Lau. 2023. The next chapter: A study of large language models in storytelling. In *Proceedings of the 16th International Natural Language Generation Conference*, pages 323–351, Prague, Czechia. Association for Computational Linguistics.

Appendix

Source	ROUGE	BLEU	METEOR	Diversity	Perplexity
În deal, pe Militari	0.35	0.07	0.75	0.78	0.73
ChatGPT	0.29	0.04	0.70	0.75	0.79
Claude	0.27	0.04	0.73	0.72	0.72
Gemini	0.27	0.03	0.74	0.68	0.77
Llama	0.03	0	0.63	0.22	0.86
Qwen	0.27	0.04	0.69	0.51	0.66
WizardLM	0.29	0.04	0.74	0.64	0.79

Table 6: Scores of the assessment metrics (pastiches for Mateiu's novel *Sub pecetea tainei*)

Source	stop.	pron.	I	art.	prep.	auxv.	adv.	conj.	neg.	verb	adj.
Mateiu	44.8	11.18	2.43	2.83	15.39	4.85	9.58	6.41	3.47	16.95	7.33
Albala	44.98	10.8	2.31	3.8	14.98	3.36	11.72	6.4	2.66	13.97	7.22
ChatGPT	45.23	12.52	3.33	5.58	13.71	4.7	7.98	4.89	1.74	15.73	7.88
Claude	44.51	10.58	1.86	3.51	14.23	7.89	8.06	5.24	2.14	17.15	7.61
Gemini	39.71	7.4	1.47	5.85	11.59	8.35	7.47	5.74	1.66	17.7	8.65
Llama	54.96	13.57	5.53	4.18	9.16	17.01	7.15	3.89	4.49	31.89	2.24
Qwen	50.96	9.09	1.66	4.35	12.02	16.99	7.7	4.25	2.53	26.47	6.56
WizardLM	45.19	14.17	0.03	4.2	14.3	2.31	9.09	5.67	2.18	14.7	5.05

Table 7: LIWC part of speech features

Source	AllPunt	Period	Comma	Question Mark	Exclamation	OtherPunct
Mateiu	23.63	5.56	11.02	0.4	0.22	6.4
Albala	20.97	3.96	12.23	0.19	0.1	4.47
ChatGPT	18.79	5.92	9.97	0.28	0.12	2.49
Claude	17.98	5.58	8.41	0.24	0	3.55
Gemini	18.11	6.48	8.58	0.52	0	2.32
Llama	16.52	7.57	6.34	0.29	0	2.32
Qwen	18.06	8.67	6.76	0.81	0.06	1.75
WizardLM	14.14	4.14	8.77	0	0	1.5

Table 8: LIWC punctuation features

Source	affect	positive	negative	female	male	insight	percept	sexual
Mateiu	6.45	3.11	3.22	0.65	1.58	2.58	3.48	0.01
Albala	6.37	3.74	2.57	1.43	1.3	2.79	3.8	0.03
ChatGPT	7.45	3.64	3.33	0.47	1.31	5.33	6.95	0
Claude	6.82	4.37	2.24	1.34	0.79	4.37	5.51	0
Gemini	9.64	4.12	5.37	0.44	0.7	7.21	4.67	0.15
Llama	5.32	2.11	2.45	0.21	1.49	9.16	7.07	0
Qwen	7.79	3.54	3.93	0.55	1.62	8.05	3.54	0
WizardLM	9.51	6.09	3.13	0.52	2.51	3.91	4.5	0

Source	past	present	future	religion	death	informal	swear
Mateiu	12.21	4.4	0.36	0.41	0.4	0.57	0.19
Albala	10.26	3.99	0.38	0.6	0.41	0.51	0.06
ChatGPT	10.4	4.83	0.72	0.5	0.09	0.28	0.06
Claude	11.78	4.37	1.45	0.28	0.24	0.34	0
Gemini	14.83	4.09	0.81	0.26	0.63	0.15	0.04
Llama	21.27	11.38	1.07	0.08	0	0.39	0
Qwen	20.23	7.34	0.97	0.1	0.45	0.58	0.03
WizardLM	11.34	3	1.3	0.2	0.1	0.07	0

Table 9: LIWC sentiment features