Pedagogika, 2025, 75(4), 355–379
http://pages.pedf.cuni.cz/pedagogika/

Out-Heroding Herod? — Author-trained GPTs and Original Works from the Perspective of Quantitative Linguistics

MICHAL MÍSTECKÝ

Abstract: Goals *– The paper compares texts created by GPT models trained on the works of prominent Czech authors and the pieces of literature they actually wrote. The goal is to find out (1) whether there are any differences between the two and, if so, (2) in what sphere of language these differences are most prominent.*

Methods *– The authors used for building the GPTs are Karel Čapek, Jaroslav Hašek, Franz Kafka, and Vladislav Vančura. The corpus contains 40 1,000-word text samples for each of them, 20 of them produced by the respective GPT and 20 taken from the original works. Two investigations are carried out – the first involves calculating 30 morphological, syntactic, and lexical markers for each text; the second is based on most-frequent-element analyses. The results of the first set are tested on statistical significance via Mann–Whitney U Test.*

Results *– The chatbots do not reflect colloquiality of style and conversational interaction very well and tend to make the texts more narrative. The best results are obtained for Karel Čapek, the worst for Franz Kafka. The stylometric analyses almost always distinguish the AI- and human-generated pieces of language.*

Conclusions *– The texts produced by the author-trained GPTs are still very well distinguishable from those produced by real writers. However, from the viewpoint of teaching practice, the chatbots may be used in critical comparisons with the original texts or ones produced by pupils/students.*

Keywords: *quantitative linguistics, stylometry, AI, chatbot, ChatGPT, Czech literature*

1. INTRODUCTION

With the development of neural reading, how AI will change the manner in which we read and receive texts has been much debated (Piorecký & Husárová, 2018), mostly in connection to reading literacy (Kosmas et al., 2025; Tao et al., 2025). Kalantzis and Cope (2025) provide a deep insight into the phenomena, comparing the advent of generative AI with the inventions of movable type and the printing press. On the grounds of these expectations, author-trained language

DOI: 10.14712/23362189.2025.4945

models – i.e. chatbots which are trained upon a large amount of data pertaining to one writer – may potentially be of great use in language and literature classes. Their greatest benefit lies in the fact that they enable pupils/students to subconsciously grasp the workings of an author's style (e.g. via generating texts similar to the ones produced by the authors they focus on), without the need to read a lot of their books or to memorise theoretical patterns. Therefore, this way of acquiring style may be attractive for schoolchildren as a result of its time-saving nature, interactivity, and joyfulness (Kosmas et al., 2025). However, what is of paramount importance is the question of whether these models really emulate the authorial styles.

The present study investigates this matter from the perspective of quantitative linguistics, the greatest assets of which are precision, the unbiased nature of the research, and the intersubjectivity of the results (Davidson, 2001). More specifically, the research focuses on how close to each other the works of a particular author and the texts produced by a language model trained on them are in terms of morphology, syntax, and lexis. Besides counting and statistically comparing various indices connected to these domains, the paper also sketches the potential of MFW/MFC[1] analyses, usually used for authorship attribution, for bringing yet another viewpoint into the game.

The paper draws inspiration from the existing studies comparing human-authored and AI-generated texts (Zaitsu & Jin, 2023; Rebora, 2023; O'Sullivan, 2024; Mikros, 2025; Milička et al., 2025). These studies share two main features: (1) they mostly use stylometric analyses, focusing on the most frequent elements (e.g. most frequent words, function word frequencies, part of speech bigrams), i.e. on deep style features, which are difficult for a human being consciously to control; (2) they arrive at the conclusion that the machine-generated texts are still very well distinguishable from the human-written ones. This study, which we believe is the first to address the topic in Czech didactics (as of May 2025), tries to overcome this narrowed viewpoint by researching concrete style markers with clear-cut linguistic interpretations. In this way, it will be possible to explain the possible differences between the two types of texts in a less black-box manner.

The paper is organised as follows. In Part 2, the design of the research, calculated indices, and the principles and parameters of the MFW/MFC analyses are outlined. Part 3 presents the processed results with accompanying interpretations. Part 4 closes the paper with a synthesis of the overlap of both the text types and spells out the limitations of the investigations; it also contains a review of the implications of the results for teaching practice.

[1] The abbreviations stand for "most frequent words" and "most frequent characters". The workings of the analyses are explained in Part 2 of the paper.

2. Research Design

In order to train the chatbots, four Czech fiction writers were selected – Karel Čapek (1890–1938), Jaroslav Hašek (1883–1923), Franz Kafka (1883–1924), and Vladislav Vančura (1890–1942). They match the following criteria: (1) they wrote mostly prose; (2) they authored novels, which means more data is available for training; (3) their texts are easily available, as they wrote most of their books in the interwar period (1918–1939); (4) their distinctive writing styles were the core of many literary studies, although their focus on language is mostly cursory (Mukařovský, 1939; Daneš, 1954; Mishra, 2023; Kundera, 1960).[2] The chatbots were trained within the GPT-4 framework (OpenAI, 2025) using the works listed in Table 1. All of the chatbots are publicly available under the commercial names CHAP.ek (= Karel Čapek), #ek (= Jaroslav Hašek), KA2F (= Franz Kafka), and VVV (= Vladislav Vančura).

These chatbots were asked to produce texts in the style of the particular author with a length of 20,000 words. The exact prompt was: "Napiš text o rozsahu 20 000 slov, který bude respektovat autorský styl XY (= např. Vladislava Vančury). Neveď konverzaci, piš text." [= Write a 20,000-word text which will respect the authorial style of XY (= e.g. of Vladislav Vančura). Do not lead a conversation, write a text.] The chatbots never produced such an amount of language within one answer; they had to be prompted to produce more texts, mostly with simple "pokračuj" (= "continue"), or by explicitly forcing them into writing more extensive chapters and keeping the chosen style. In the event that two variants of the text were offered by the GPT, the first one was always included in the corpus. The results differ in terms of genre and text types, namely:

a) CHAP.ek produced a text that it called O obyčejném člověku (= On an Ordinary Person). It tended to close the story before the limit of the words was reached; when prompted to produce more text, it started to vary the narrative genre-wise, adding short stories and essays revolving around the protagonist of the main text, Mr Kubeš.

b) #ek produced a novel about the adventures of a Mr Pěšina (= Pathway, possibly a nomen omen of the picaresque character). It used quite elaborate chapter names, which were kept as part of the corpus (in the case of the texts produced by the other chatbots, the chapter names were excluded, as they contained numbers only).

c) KA2F did not give its work any name; it produced a complicated story with a visibly anti-bureaucratic bent. When the word limit was reached, it was evident that the GPT wanted to develop the story further.

d) VVV wrote a novel/novella (it uses both these genre labels throughout the

[2] The texts by Franz Kafka that were used to train the respective chatbot were translations from German. This fact does not seem to impede the analysis, but it is noteworthy, as the conclusions formulated here in this regard may pertain not only to Kafka himself, but to the translator as well.

Table 1. Overview of the training data used to create the GPTs. The English translations provided were those under which the works were published in English-speaking countries, or those that were closest to the original title. As for Kafka's posthumously published works, they were ordered according to the time of their writing.

Chatbot name	Author	Training data	Training data [ENG]	Training data size [words]
CHAP.ek	Karel Čapek	Krakatit (1924)	Krakatit	291,404
		Hordubal (1933)	Hordubal	
		Povětroň (1934)	Meteor	
		Obyčejný život (1934)	An Ordinary Life	
		Válka s Mloky (1936)	War with the Newts	
#ek	Jaroslav Hašek	Osudy dobrého vojáka Švejka za světové války (1921–1923, three volumes + the unfinished fourth)	The Good Soldier Švejk	344,434
KA2F	Franz Kafka	Amerika (1912)	Amerika	315,422
		Rozjímání (1913)	Contemplation	
		Ortel (1913)	The Judgment	
		Proces (1914–1915)	The Trial	
		Proměna (1915)	The Metamorphosis	
		Dopis otci (1918)	Letter to His Father	
		V kárném táboře (1919)	The Penal Colony	
		Venkovský lékař (1920)	A Country Doctor	
		Umělec v hladovění (1922)	A Hunger Artist	
		Zámek (1922)	The Castle	
VVV	Vladislav Vančura	Pekař Jan Marhoul (1924)	Baker Jan Marhoul	120,777
		Pole orná a válečná (1925)	Ploughshares into Swords	
		Rozmarné léto (1926)	Summer of Caprice	
		Markéta Lazarová (1931)	Markéta Lazarová	

chat) called Poutníci času (= Wanderers of Time), but closes the story multiple times before the final word limit is reached; it is thus pushed to produce more chapters, which are loosely linked to the original novel/novella.

These 20,000-word texts produced by the individual GPTs were automatically split into 1,000-word chunks, which were used in the subsequent analysis. The chat-produced texts thus totalled 80 (= 20 for each author's GPT). As their counterparts, the original training data was also split into 1,000-word chunks, out of which 20 were randomly selected for each author. This design was chosen as it is assumed that the GPT takes into account all of the particular author's oeuvre when creating new texts. The final corpus thus includes 160 texts. Because of the idiosyncratic use of punctuation, there is a certain fluctuation in the text sizes (994–1,007 words).

Next, 30 indicators/indexes/markers were counted for each text of the corpus; these markers cover the areas of morphology [23 – see a) to h) below], syntax [2 – see i) and j) below] and lexis [5 – see k) to p) below], and were selected as they have already proved their utility in discriminating between styles (e.g. Kubát, 2016; Cvrček et al., 2020a; Místecký & Radková, 2020; Místecký & Melka, 2021). They include:

a) the relative frequencies of parts of speech (abbreviations: noun – subst, adjective – adj, pronoun – pron, numeral – num, verb, adverb – adv, preposition – prep, conjunction – conj, particle – part, interjections – interj; for the workings, see Vondráček, 2013);

b) the relative frequencies of grammatical cases (abbreviations: nominative – nom, genitive – gen, dative – dat, accusative – acc, vocative – voc, locative – loc, and instrumental – inst; for the workings, see Janda et al., 2022);

c) singularity (sg), which is counted as the ratio of singular grammatical forms to plural and dual (which is present only residually in Czech) ones;

d) perfectivity (perf), calculated as the ratio of perfective verbs to imperfective and biaspectual ones (for the workings of aspect, see Dahl, 2000);

e) the relative frequency of the past participle (pastpart), which is the main indicator of the past tense in the Czech language;

f) the relative frequency of the perfect participle (perfpart), which is the main component of the Czech descriptive passive;

g) the ratio of the frequency of deverbative adjectives and the total of adjectives in the text (dev_adj);

h) the moving-average morphological richness (mamr_100), which amounts to the subtraction of the mattr value of a non-lemmatised text from the mattr value of a lemmatised one [for mattr, see k)];

i) the ratio of the subordinating conjunctions in the text and the sum of the subordinating and coordinating conjunctions (sub);

j) verb distances (vd), which are calculated as the number of words to be found in between two verbal forms divided by the number of distances;

k) mattr_100 (moving-average type–token ratio), which measures the lexical diversity of a text, taking into account the amount of word repetition within a linguist-defined window (in our research set to 100 words; Covington & McFall, 2010);

l) atl (average token length), which calculates the average length of a word counted in letters;

m) the ratio of hapax legomena (= the words that occur only once in a text) to the total of the words in the text (hapax);

n) thematic concentration (tc), which indicates the amount of concentration of the text on topics expressed by autosemantic words (for the workings, see Kubát, 2016);

o) secondary thematic concentration (stc), which works in the same way as tc, but takes into account more words from the text (for the workings, see Kubát, 2016).

As concerns the indices k), m), n), and o), the texts that entered the analysis in these cases were lemmatised, i.e. all the wordforms in them were replaced with their base forms (e.g. the sentence "the dogs were running towards the gates" would be lemmatised as "the dog be run towards the gate"). This solution was chosen as these indices focus on the topics which are represented by the base forms and not on grammar (the words "women" and "woman" thus count as one lemma = "woman"). Concerning the relativisation of values in a), e), and f), punctuation is disregarded. The results for a) – g) and for j) were obtained using the MorphoDita web application (Straková et al., 2014); the remaining markers were counted via the QuitaUp application (Cvrček et al., 2020b).

The comparison involved the original and GPT-generated texts for each author. The texts were coded in the following way: CHAT/REAL_author_number (e.g. CHAT_capek_01, REAL_vancura_16). The comparisons were carried out for each index separately. The statistical significance (or insignificance) of the difference in the values of an index was determined on the grounds of Mann–Whitney U Test (Mann & Whitney, 1947). The level of significance was set at 0.05.

Finally, most-frequent-element analyses were conducted (Eder et al., 2016). In this case, the main outcomes are dendrograms clustering texts that belong together according to the selected criteria. The distances between the texts were calculated on the grounds of Classic Delta, which is highly recommended in authorship attribution studies (Burrows, 2002). The elements analysed were (1) the 100 most frequent wordforms; (2) the 100 most frequent character bigrams; (3) the 50 most frequent character bigrams. Character bigrams correspond to pairs of characters that are the products of the window-like splitting of a text; for instance, the sentence "dogs bark" contains the following character bigrams: "do", "og", "gs", "s_", "_b", "ba", "ar", "rk". In this way, the conscious choice of elements which may have an impact on the results of (1) is inhibited, as it is highly improbable that an author may have any control over the production of character bigrams.

3. Results

The results for the statistical testing of the indices are listed in Tables 2–5. Each author is discussed separately, and the indices are interpreted in accordance with their presentation in Part 2.

Regarding the Karel Čapek corpus, the chatbot underestimated the author's use of pronouns, particles, and interjections, while placing more emphasis on adverbs, prepositions, and conjunctions. This indicates that it stressed the meaning-infused part of Čapek's books and exaggerated the complexity of his noun phrases (= prepositions) and sentences (= conjunctions; this contradicts the role of commas in Čapek's books, as emphasised in Mukařovský, 1939); however, the texts it produces lack Čapek's command of common speech, which is manifested in his employment of pronouns, particles, and interjections. The virtual absence of the vocative in the GPT-generated texts is in line with this tendency, whereas the elevated use of subordination reflects the chatbot's inclination towards intellectualisation. Nevertheless, the more frequent use of the accusative, locatives, and past participles may point at the narrativisation of Čapek's style, as stories usually necessitate the past tense, confrontation, and spatiotemporal localisation. The GPT also stresses individual agency, as it utilises statistically significantly fewer past participles than Čapek himself did. Finally, the chatbot does not successfully imitate the range of the original author's word-stock, as the number of hapaxes in the texts it generated is statistically significantly lower than in the original ones. This may be ascribed to the high frequencies of prepositions and conjunctions, which, as functional words, tend to be repeated often in texts.

To conclude, the Čapek-trained GPT fabricated linguistically complicated stories with more repetition and less contact with common speech; the products are thus, in a way, more stylised than the original. On the other hand, 14 out of 30 indices manifested statistically significant differences between the two sets of texts, this meaning that the GPT performed quite well for most of the markers.

As for the Jaroslav Hašek corpus, there are several features shared with that of Karel Čapek. First and foremost, the chatbot tends to limit colloquiality by means of suppressing pronouns (on the role of the relative ones in the Švejk books, see Daneš, 1954), particles, and interjections, and manifests certain traits of rich language, which may reflect the multifaceted character of Hašek's adventure stories (the higher frequencies of nouns, and elevated values of mattr_100 and mamr_100). Moreover, the interactional character of Hašek's language is lowered, too, which is indicated by the backgrounding of the dative (= the primary case of the indirect object in speeches) and the instrumental (which is governed by the "with" preposition, e.g. in "talk with sb"). Another case that is reduced in frequency is the locative – spatiotemporal orientation may thus be of less importance to the GPT, which is possibly connected to the fact that the

Table 2. The overall results for the indices counted on the Karel Čapek corpus. For the abbreviations, see Part 2. The means were rounded to the nearest hundredth, the p-values to the nearest hundred thousandth. As for statistical testing and significance, U_stat = the Mann–Whitney U Test statistic, Y = the difference is significant (p_value < 0.05), N = the difference is not significant (p_value > 0.05).

	mean_chat	**mean_real**	**U_stat**	**p_value**	**significance**
pastpart	0.1176	0.0595	398	0.0000	Y
part	0.0179	0.0595	3	0.0000	Y
voc	0.0011	0.0242	38	0.0000	Y
interj	0.0001	0.004	61	0.0000	Y
sub	0.4595	0.3673	325	0.0008	Y
conj	0.1101	0.0939	322	0.0010	Y
acc	0.3095	0.2714	307	0.0040	Y
hapax	0.3139	0.3462	100.5	0.0074	Y
sg	0.8186	0.7532	299	0.0077	Y
prep	0.0865	0.0776	295	0.0106	Y
pron	0.14	0.1576	112	0.0179	Y
adv	0.0918	0.0823	284.5	0.0231	Y
perfpart	0.002	0.0049	117	0.0256	Y
loc	0.0751	0.0646	281	0.0294	Y
dev_adj	0.0091	0.0273	142	0.1101	N
stc	0.0313	0.0546	143	0.1264	N
atl	4.5208	4.6918	146.5	0.1516	N
num	0.013	0.012	253	0.1555	N
dat	0.0526	0.0602	151	0.1895	N
vd	4.69	5.4361	160	0.2853	N
mattr_100	0.744	0.9233	237.5	0.3167	N
tc	0.0107	0.0322	163	0.3169	N
inst	0.0629	0.0713	167	0.3793	N
mamr_100	0.0855	0.0807	232.5	0.3867	N
perf	0.2964	0.2857	229	0.4408	N
verb	0.2251	0.2206	225	0.5075	N
adj	0.0714	0.0844	175	0.5075	N
gen	0.1268	0.1334	178	0.5609	N
subst	0.2441	0.2454	219	0.6168	N
nom	0.372	0.375	195	0.9031	N

Table 3. The overall results for the indices counted on the Jaroslav Hašek corpus. For the abbreviations, see Part 2. The means were rounded to the nearest hundredth, the p-values to the nearest hundred thousandth. As for statistical testing and significance, U_stat = the Mann–Whitney U Test statistic, Y = the difference is significant (p_value < 0.05), N = the difference is not significant (p_value > 0.05).

	mean_chat	**mean_real**	**U_stat**	**p_value**	**significance**
voc	0.0008	0.0177	14.5	0.0000	Y
nom	0.4022	0.299	377	0.0000	Y
num	0.0353	0.0149	376	0.0000	Y
dat	0.0485	0.0774	28	0.0000	Y
vd	4.19	5.3624	49	0.0000	Y
prep	0.0865	0.1093	51	0.0001	Y
sub	0.3986	0.4896	51.5	0.0001	Y
pron	0.1301	0.154	52	0.0001	Y
part	0.0075	0.0139	59	0.0001	Y
gen	0.1231	0.1646	59	0.0001	Y
mattr_100	0.7663	0.7341	338.5	0.0002	Y
interj	0.0003	0.0019	107.5	0.0028	Y
verb	0.2241	0.2044	310	0.0031	Y
mamr_100	0.0696	0.0816	91	0.0033	Y
pastpart	0.0921	0.0811	301	0.0066	Y
sg	0.7749	0.807	100	0.0071	Y
loc	0.0693	0.0849	105	0.0106	Y
inst	0.0593	0.0731	122	0.0360	Y
subst	0.2695	0.2555	276	0.0411	Y
stc	0.0692	0.0494	267	0.0720	N
perfpart	0.0029	0.0039	134	0.0764	N
dev_adj	0.0099	0.0171	147.5	0.1327	N
tc	0.0462	0.0283	246	0.2148	N
acc	0.2968	0.2833	241	0.2733	N
perf	0.3942	0.3726	240	0.2853	N
hapax	0.352	0.3347	236	0.3366	N
atl	4.6649	4.6224	236	0.3369	N
conj	0.097	0.0937	227	0.4735	N
adj	0.0708	0.0747	181	0.6168	N
adv	0.079	0.0778	213	0.7353	N

protagonist of the GPT-created story, Mr Pěšina, stays in one place for most of the novel (in comparison to Josef Švejk, who is constantly moving / being sent somewhere). The differences between the two stories are also manifested in the reduced amount of singularity in the chatbot-fabricated texts – whilst Mr Pěšina is an active person who organises many (mock) social events for the people from the village he lives in, Josef Švejk mostly focuses on himself, engages in one-to-one conversations, and does not address masses of people. This marker thus shows a crucial distinction between how the two stories are built.

Next, the GPT endeavours to downplay the intellectualisation features, e.g. by diminishing the occurrences of prepositions, genitives, and subordination, by lowering the verb distances, and by raising the number of verbs, past participles, and conjunctions (thus replacing noun phrases with verb phrases; the verb-oriented style was already declared to be typical of Švejk by Daneš, 1954).

Last but not least, the frequent occurrences of the nominative and numerals are mostly linked to the fact that the chatbot divided its literary production into chapters headed by lengthy titles (in the form of "Chapter 1, in which […]"). This peculiar feature reflects the practice of authors of picaresque novels, which was, however, not employed by Jaroslav Hašek. Overall, the statistically significant differences were shown for 19 out of 30 indices, which makes the style imitation task result less successful than in the case of Čapek.

The results of the analysis of the Franz Kafka corpus manifest several contradictory tendencies. On the one hand, the GPT is prone to exaggerate the occurrences of verbs and past participles and downplay those of prepositions and subordination, which reveals a tendency towards direct storytelling; on the other hand, however, it diminishes the frequencies of conjunctions, adverbials, which add circumstances to the actions that are described, of particles, interjections and vocatives, which connect the story with the real world, and of perfective verbs, enhancing the verb distances, average token length, frequencies of nouns, and use of the perfect participle, possibly with the intention of showing the bureaucracy-mocking language Kafka used in his novels. This second trend appears to be stronger and is corroborated by other findings. For instance, the considerable reductions in the frequency of the dative and the accusative point at the story going on without much interaction among the characters and at the elevated employment of objectless verbs. The substantially increased frequencies of the nominative further foster this "encapsulation" of action and provide us with a character who is ostensibly active, but whose actions do not lead anywhere. This last point is also confirmed by the reduction in the number of instances of the locative case found in the chatbot-produced texts.

Quite a specific situation emerged with regard to the lexical indices. The texts produced by the Kafka-taught GPT are less lexically diverse than the original ones (= they

Table 4. The overall results for the indices counted on the Franz Kafka corpus. For the abbreviations, see Part 2. The means were rounded to the nearest hundredth, the p-values to the nearest hundred thousandth. As for statistical testing and significance, U_stat = the Mann–Whitney U Test statistic, Y = the difference is significant (p_value < 0.05), N = the difference is not significant (p_value > 0.05).

	mean_chat	**mean_real**	**U_stat**	**p_value**	**significance**
verb	0.2818	0.2235	384	0.0000	Y
adv	0.071	0.1222	1	0.0000	Y
nom	0.4293	0.299	400	0.0000	Y
dat	0.0493	0.0883	18	0.0000	Y
acc	0.2398	0.3014	38	0.0000	Y
pastpart	0.1132	0.0748	355	0.0000	Y
perfpart	0.0181	0.0052	396	0.0000	Y
stc	0.0096	0.0505	4	0.0000	Y
tc	0.0008	0.0328	60	0.0000	Y
conj	0.091	0.1066	58	0.0001	Y
mamr_100	0.1094	0.0899	334	0.0003	Y
vd	5.5791	4.9777	323.5	0.0009	Y
prep	0.0764	0.0891	83	0.0016	Y
mattr_100	0.7034	0.7318	86.5	0.0022	Y
subst	0.208	0.185	312	0.0026	Y
sub	0.375	0.4344	93	0.0040	Y
loc	0.069	0.0854	94	0.0043	Y
hapax	0.2614	0.3128	102	0.0083	Y
interj	0	0.0005	140	0.0096	Y
voc	0.0007	0.0058	130	0.0101	Y
atl	4.86	4.44	292	0.0133	Y
perf	0.3129	0.3624	112	0.0179	Y
part	0.0201	0.0248	126.5	0.0482	Y
dev_adj	0.0173	0.0328	138.5	0.0954	N
inst	0.0851	0.0922	155.5	0.2339	N
sg	0.8286	0.8112	243	0.2503	N
adj	0.0749	0.0687	242.5	0.2559	N
num	0.0125	0.01	235.5	0.3437	N
pron	0.1643	0.1696	175	0.5075	N
gen	0.1268	0.128	187	0.7353	N

manifest lower mattr_100 values and a lower number of hapaxes), but they tend to be less thematically concentrated (according to both tc and stc) as well. This may be ascribed to the fact that whilst the chatbot repeats mostly synsemantic expressions, which do not count as thematic words (but which may raise the level of morphological richness – mamr_100), Kafka's fiction tends to elaborate lengthily on various subjects which are represented by autosemantic words (nouns, adjectives, verbs). All in all, the success rate of the Kafka chatbot is rather low, with up to 23 out of 30 markers manifesting statistically significant differences between the two text groups.

Finally, the texts created by the GPT trained on Vladislav Vančura's works seem to manifest similar inconsistencies to those mentioned above. Again, the colloquial and dialogical nature of the original is not properly reflected in the GPT products, as these manifest lower frequencies of interjections, datives, and vocatives. Moreover, the imitation of the static and intentionally old-fashioned language typical of Vančura fell below expectations, too, as the chatbot-crafted novel/novella lowers the frequencies of adjectives, genitives, perfect participles, deverbative adjectives, and hapaxes, and the amount of lexical diversity (mattr_100; for the growing richness of Vančura's lexis, see Kundera, 1960); it also makes use of shorter verb distances. On the other hand, the GPT seems to try to compensate for these drawbacks by boosting morphological diversity (mamr_100), and utilising more subordination, particles (e.g. "despite", "perhaps"), and longer words. The raised frequencies of adverbs, locatives, perfective verbs, and part participles reveal an inclination towards narrativisation, which links these texts with those generated by the Čapek chatbot.

There are two more indices that draw the texts that were produced away from the style of Vančura – first, the chatbot uses more singular words, which is not in line with the sort of collectivist approach that links Vančura's books with chronicles; second, the texts that are generated are more thematically concentrated (from the perspective of stc), which also points at a more detailed treatment of the elements of the story, which contradicts the sketchy breadth of narration one associates with Vančura.

The summary presented in Table 6 further confirms the tendencies touched upon in the previous interpretations. The chatbots tend to downplay the features connected to everyday spoken language (interjections, vocatives, particles – except for Vančura, datives – except for Čapek) and stylistic specificities (lowered number of hapaxes – except for Hašek), while they strengthen, to a lower extent, the narrative traits of texts (past participles, adverbs – except for Hašek and Kafka). In the case of subordination and the locative, the GPTs seem to follow logical inductions that finally prove to be wrong – they speculate that intellectual and story-focused writers such as Čapek and Vančura will use more subordination structures and locatives, whereas Hašek will be more straightforward and limited to one particular place

Table 5. The overall results for the indices counted on the Vladislav Vančura corpus. For the abbreviations, see Part 2. The means were rounded to the nearest hundredth, the p-values to the nearest hundred thousandth. As for statistical testing and significance, U_stat = the Mann–Whitney U Test statistic, Y = the difference is significant (p_value < 0.05), N = the difference is not significant (p_value > 0.05).

	mean_chat	**mean_real**	**U_stat**	**p_value**	**significance**
interj	0.0002	0.0034	22	0.0000	Y
dat	0.0342	0.0719	20	0.0000	Y
pastpart	0.1236	0.0862	383	0.0000	Y
atl	4.91	4.24	0	0.0000	Y
hapax	0.2842	0.4133	0	0.0000	Y
vd	4.0583	4.8186	44	0.0000	Y
adj	0.0574	0.0777	58.5	0.0001	Y
part	0.0181	0.0121	346	0.0001	Y
voc	0.0022	0.008	54	0.0001	Y
loc	0.077	0.0602	334	0.0003	Y
mamr_100	0.0844	0.0695	327.5	0.0006	Y
dev_adj	0.0116	0.0364	75.5	0.0006	Y
mattr_100	0.7467	0.7772	73	0.0006	Y
adv	0.0738	0.0602	326.5	0.0007	Y
stc	0.0373	0.0203	325	0.0008	Y
gen	0.1321	0.1706	82	0.0015	Y
sg	0.7809	0.7402	304	0.0051	Y
sub	0.3861	0.3344	286	0.0207	Y
perf	0.3606	0.3296	283	0.0256	Y
perfpart	0.0046	0.007	118	0.0275	Y
nom	0.6802	0.6714	267	0.0720	N
subst	0.2547	0.2688	133.5	0.0742	N
verb	0.2429	0.2331	263	0.0908	N
pron	0.1514	0.1462	242.5	0.2558	N
tc	0.0137	0.0084	234.5	0.3347	N
num	0.0131	0.0112	233	0.3791	N
acc	0.278	0.2869	172	0.4570	N
conj	0.1026	0.0999	227	0.4734	N
inst	0.0748	0.0727	221	0.5792	N
prep	0.0859	0.0873	180.5	0.6072	N

Table 6. Summary of the results from the perspective of statistical significance. Y = the difference is significant for the given index, N = the difference is not significant, Y_count = the total number of the authorial comparisons in which the given index is significant (max = 4); ↑ = the mean of the given index in the chat-produced texts is higher than that in the original texts, ↓ = the mean of the given index in the chat-produced texts is lower than that in the original texts. It is to be noted that all the indices produced statistically significant difference for at least one authorial comparison.

index	Y_count	Čapek	Hašek	Kafka	Vančura
interj	4	Y (↓)	Y (↓)	Y (↓)	Y (↓)
loc	4	Y (↑)	Y (↓)	Y (↓)	Y (↑)
part	4	Y (↓)	Y (↓)	Y (↓)	Y (↑)
pastpart	4	Y (↑)	Y (↑)	Y (↑)	Y (↑)
sub	4	Y (↑)	Y (↓)	Y (↓)	Y (↑)
voc	4	Y (↓)	Y (↓)	Y (↓)	Y (↓)
adv	3	Y (↑)	N	Y (↓)	Y (↑)
dat	3	N	Y (↓)	Y (↓)	Y (↓)
hapax	3	Y (↓)	N	Y (↓)	Y (↓)
mamr_100	3	N	Y (↓)	Y (↑)	Y (↑)
mattr_100	3	N	Y (↑)	Y (↓)	Y (↓)
perfpart	3	Y (↓)	N	Y (↑)	Y (↓)
prep	3	Y (↑)	Y (↓)	Y (↓)	N
sg	3	Y (↑)	Y (↓)	N	Y (↑)
vd	3	N	Y (↓)	Y (↑)	Y (↓)
acc	2	Y (↑)	N	Y (↓)	N
atl	2	N	N	Y (↑)	Y (↑)
conj	2	Y (↑)	N	Y (↓)	N
gen	2	N	Y (↓)	N	Y (↓)
nom	2	N	Y (↑)	Y (↑)	N
perf	2	N	N	Y (↓)	Y (↑)
pron	2	Y (↓)	Y (↓)	N	N
stc	2	N	N	Y (↓)	Y (↑)
subst	2	N	Y (↑)	Y (↑)	N
verb	2	N	Y (↑)	Y (↑)	N
adj	1	N	N	N	Y (↓)
dev_adj	1	N	N	N	Y (↓)
inst	1	N	Y (↓)	N	N
num	1	N	Y (↑)	N	N
tc	1	N	N	Y (↓)	N

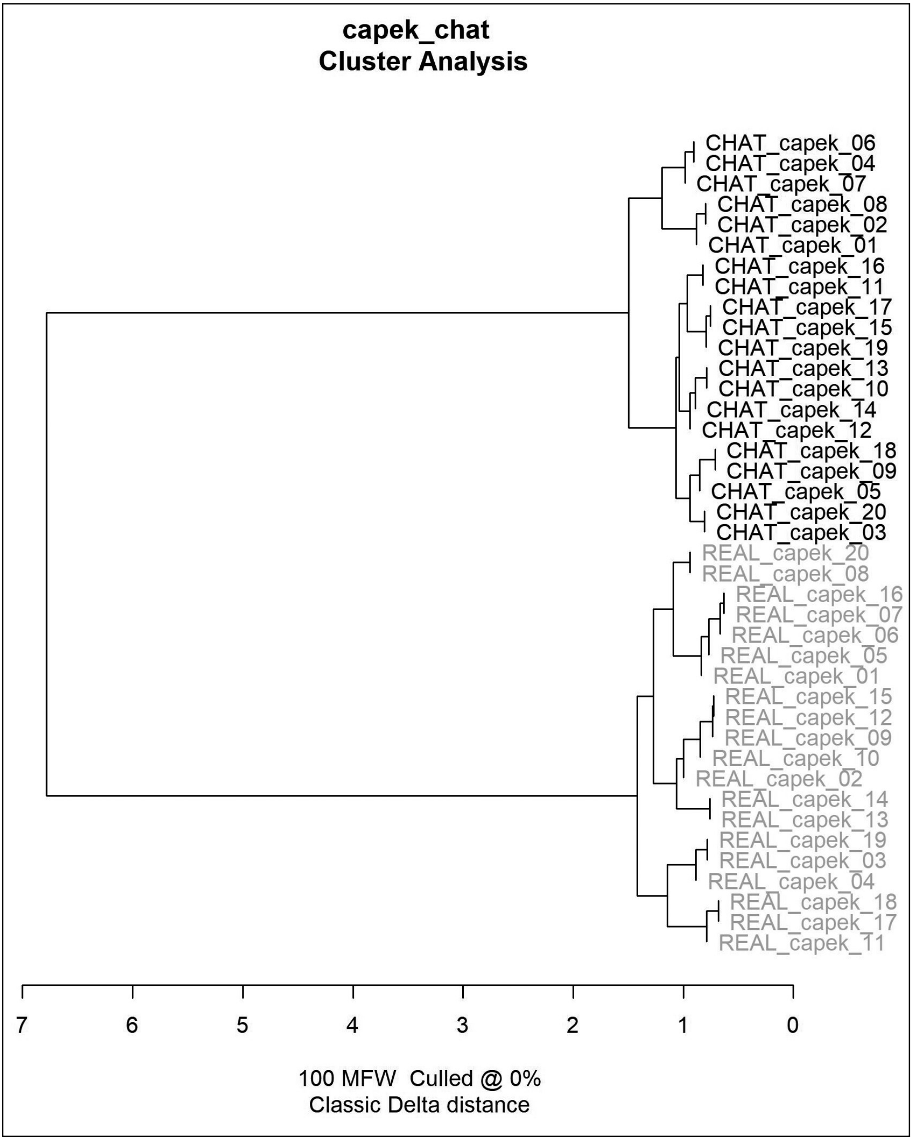

Figure 1. Cluster analysis of the Karel Čapek corpus on the grounds of the 100 most frequent words and the Classic Delta distance.

(possibly the army) and Kafka more abstract, with less attention paid to spatio-temporal relations. The same tendency – using information that is not inherent in the training tests, but stems from certain general knowledge the GPTs have about the authors being researched – may explain, for instance, the frequency of the perfect participle and employment of high verb distances in Kafka's texts (= traits of administrative style), or the use of more singulars and prepositions in Čapek's products (= his tendency to individualism and nominal structures). The theoretical knowledge of ChatGPT may thus have played a role in its stylistic misalignments.

Last but not least, the results of the most-frequent-element analysis are presented. In total, 12 analyses were performed (three types and four comparisons – see Part 2); out of these, nine classified the chat-created and original texts into two distinctly separated clusters. For the sake of illustration, one such clustering is presented in Figure 1; part of the frequency data it is based on is listed in Table 7. It is discernible that e.g. the chat-generated texts underrate the frequency of the "a" conjunction, which is the cornerstone of Čapek's coordination-driven style. The GPT also does not reflect the importance of the "si" pronoun for the author, as well as his liking for the "tak" particle.

Three analyses provide a slightly different classification picture from the others; these are: (1) the analysis of the 50 most frequent bigrams in the Karel Čapek corpus; (2) the analysis of the 100 most frequent bigrams in the Jaroslav Hašek corpus, and (3) the analysis of the 50 most frequent bigrams in the Jaroslav Hašek corpus (see Figures 2–4). Three situations are encountered: in (1), one real text (REAL_capek_14) penetrated among the chat-produced ones; in (2), two chat-produced texts (CHAT_hasek_01 and CHAT_hasek_02) ended up being classified among the real ones; in (3), the two-texts from (2) were rather loosely connected to the GPT-produced cluster, but were closely linked to one real Hašek's text (REAL_hasek_04). All the specific texts are presented as examples 1–4 below; the original texts are provided as footnotes; the translations are by the author of the paper, who was assisted by DeepL (DeepL Translator Team, 2025).

The passage from Čapek, which was incorrectly labelled as a product of the chatbot, contains quite a long philosophical contemplation on the nature of memory and its loss, which, in its essay-like manner, harmonises with how the chatbot conceived Čapek's style. On the other hand, the reason behind the grouping of REAL_hasek_04 with the chat-produced texts is more down-to-earth: it was probably due to the prominent use of numerals, which is typical of the chapter names of the AI-fabricated Hašek texts, as discussed above.

The only palpable "success" of the chatbots in these analyses may be the classification of two texts that endeavour to emulate Hašek (CHAT_hasek_01 and CHAT_hasek_02) among his real pieces of language. It is symptomatic that these are the first two texts generated by the #ek GPT, as at the beginning, it may have

Table 7. The first 20 wordforms, their English translations, and relative frequencies in the first three texts by the Čapek chatbot (CHAT) and by Karel Čapek (REAL). Relative frequencies are counted as the ratio of the absolute frequency of a wordform and the total of the wordforms in the text, multiplied by 100. For the abbreviations, see Part 2; rel = relative pronoun.

word-form	wordform [ENG]	CHAT			REAL		
		capek_01	capek_02	capek_03	capek_01	capek_02	capek_03
a	and	3.40	3.50	2.51	3.00	5.00	6.59
se	oneself [acc]	2.10	3.00	3.31	3.20	3.50	2.20
to	it	1.00	1.50	0.90	0.80	2.30	3.90
na	on	1.30	1.70	1.30	1.20	1.30	1.70
ale	but	1.60	1.20	1.60	2.60	0.70	0.90
v	in	1.90	2.30	1.80	0.90	1.20	1.20
je	[he/she/it] is	1.10	0.60	0.90	0.50	0.80	2.00
že	that [conj]	0.90	1.20	1.30	1.30	1.20	0.40
si	oneself [dat]	1.20	1.70	1.30	0.90	0.50	0.70
jako	like/as	1.40	1.50	1.10	1.40	0.90	0.90
když	when	0.70	1.00	0.90	1.00	1.00	0.50
s	with	1.10	1.10	0.90	0.60	0.70	1.10
by	would	0.90	0.60	0.80	0.50	0.20	0.80
do	to	0.80	0.90	0.60	0.60	0.80	0.60
byl	[he] was	1.00	0.60	1.00	0.40	1.10	0.40
ne	no	0.80	0.70	0.60	1.20	0.00	0.10
tak	that way / so	0.30	0.10	0.30	0.50	0.30	0.50
co	what / that [rel]	0.40	0.10	0.40	0.40	0.10	0.70
jen	only	0.30	0.70	0.80	0.70	0.00	0.70
z	from	0.80	0.50	0.50	0.40	0.40	0.30

been more meticulous in respecting the prescribed authorial style. The two texts feel Hašek-like, use situational humour, and satirise the bureaucratic structure of Austria-Hungary. On the other hand, the second text abounds in surreal elements, which may be seen as the first drift of the chatbot away from Hašek's style.

[1] A person who has lost their memory is like a person who has lost consciousness; even if their brain continues to be clear and normal, it is as if they have left

the ground of reality and lived outside it; bear in mind that without memory, there would be no reality for us. As a doctor, you would surely appreciate that our case of memory loss follows acute alcoholic poisoning and the physical shock caused by that nightly episode.[3] (REAL_capek_14)

[2] Overwhelmed by the many inventions of the Austrian Ministry of the Interior, Constable Flanderka had an enormous amount of unfinished business and answered the questionnaires in the stereotypical way that everything was fine with him and that his loyalty among the local population was on the Ia scale. The Austrian Ministry of the Interior invented the following scales for loyalty and steadfastness to the empire: Ia, Ib, Ic–IIa, IIb, IIc–IIIa, IIIb, IIIc–IVa, IVb, IVc. This last Roman numeral four meant, in conjunction with a, treason and the rope, b, intern, c, observe and incarcerate.[4] (REAL_hasek_04)

[3] Pěšina packed his belongings into his backpack, which took the form of an old potato sack, and headed towards the nearest large town, where, as he had heard, people wore shoes even in the middle of the week. He had with him some bread, onion, and a book called "One Hundred Ways to Look Important". He arrived at the town, where a group of homeless people had taken up residence at the railway station, and they immediately took him for competition. A city official detained him for looking as if he wanted to enter without intent, which was in violation of Ordinance 88b, regarding the intentional movement of citizens.[5] (CHAT_hasek_01)

[4] On his way to Nymburk for a meeting of anonymous receipt collectors, Pěšina wandered into a compartment that apparently existed outside ordinary reality. There was a man in a top hat feeding a fox with small croissants and a lady knitting a scarf made of beeswax. 'Welcome to the carriage of parallel lives,' said the conductor, who had a tuning fork instead of a whistle.[6] (CHAT_hasek_02)

[3] Člověk, který ztratil paměť, se podobá člověku, který ztratil vědomí; i kdyby jeho mozek nadále byl jasný a normální, je to, jako by opustil půdu skutečnosti a žil mimo ni; vězte, že bez paměti by pro nás nebylo ani skutečnosti. Zajisté jste jako doktor ocenil, že náš případ ztráty paměti následuje po akutní otravě alkoholické a po fyzickém šoku, způsobeném onou noční příhodou.

[4] Zaplaven tou spoustou vynálezů rakouského ministerstva vnitra, strážmistr Flanderka měl ohromnou spoustu restů a dotazníky zodpovídal stereotypně, že je u něho všechno v pořádku a loajalita že je mezi místním obyvatelstvem stupnice Ia. Rakouské ministerstvo vnitra vynalezlo pro loajalitu a neochvějnost k mocnářství tyto stupnice: Ia, Ib, Ic – IIa, IIb, IIc – IIIa, IIIb, IIIc – IVa, IVb, IVc. Tahle poslední římská čtverka znamenala ve spojení s a velezrádce a provaz, s b internovat, s c pozorovat a zavřít.

[5] Pěšina si sbalil věci do rance, který měl podobu starého pytle po bramborách, a vydal se směrem k nejbližšímu velkoměstu, o kterém slyšel, že tam lidé nosí boty i ve středu týdne. Měl s sebou chleba, cibuli a knihu s názvem „Sto způsobů, jak se tvářit důležitě". Dorazil k městu, kde se na nádraží usídlila skupinka bezdomovců, kteří ho ihned pokládali za konkurenci. Městský strážník ho zadržel za to, že se tvářil, jako by chtěl vstoupit bez záměru, což bylo v rozporu s vyhláškou číslo 88b o záměrném pohybu občanů.

[6] Cestou do Nymburka na setkání anonymních sběratelů účtenek Pěšina zabloudil do kupé, které patrně existovalo mimo běžnou realitu. Seděl tam muž v cylindru, co krmil lišku loupáčky, a dáma, která pletla šálu z včelích vosků. „Vítejte ve vagónu paralelních životů," pravil průvodčí, který měl místo pískací píšťalky ladící vidlici.

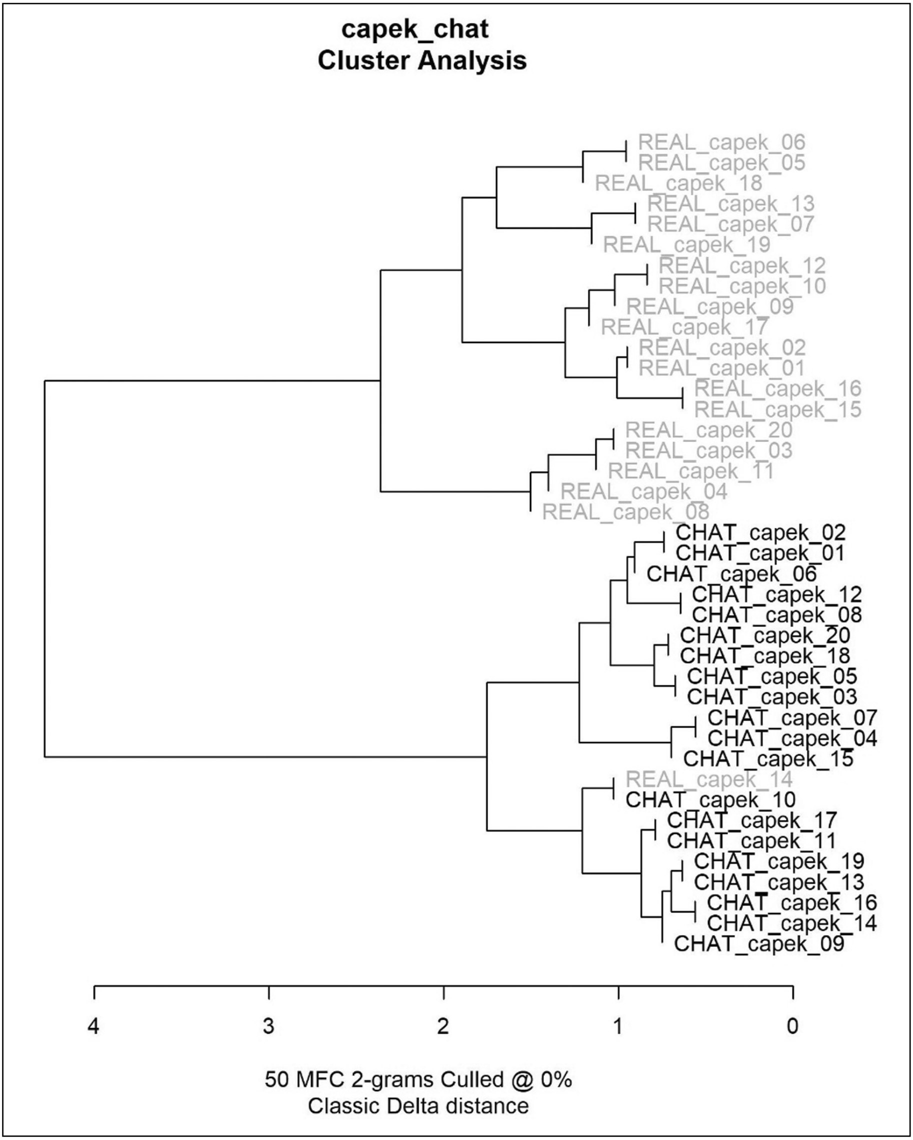

Figure 2. Cluster analysis of the Karel Čapek corpus on the grounds of the 50 most frequent character bigrams and the Classic Delta distance.

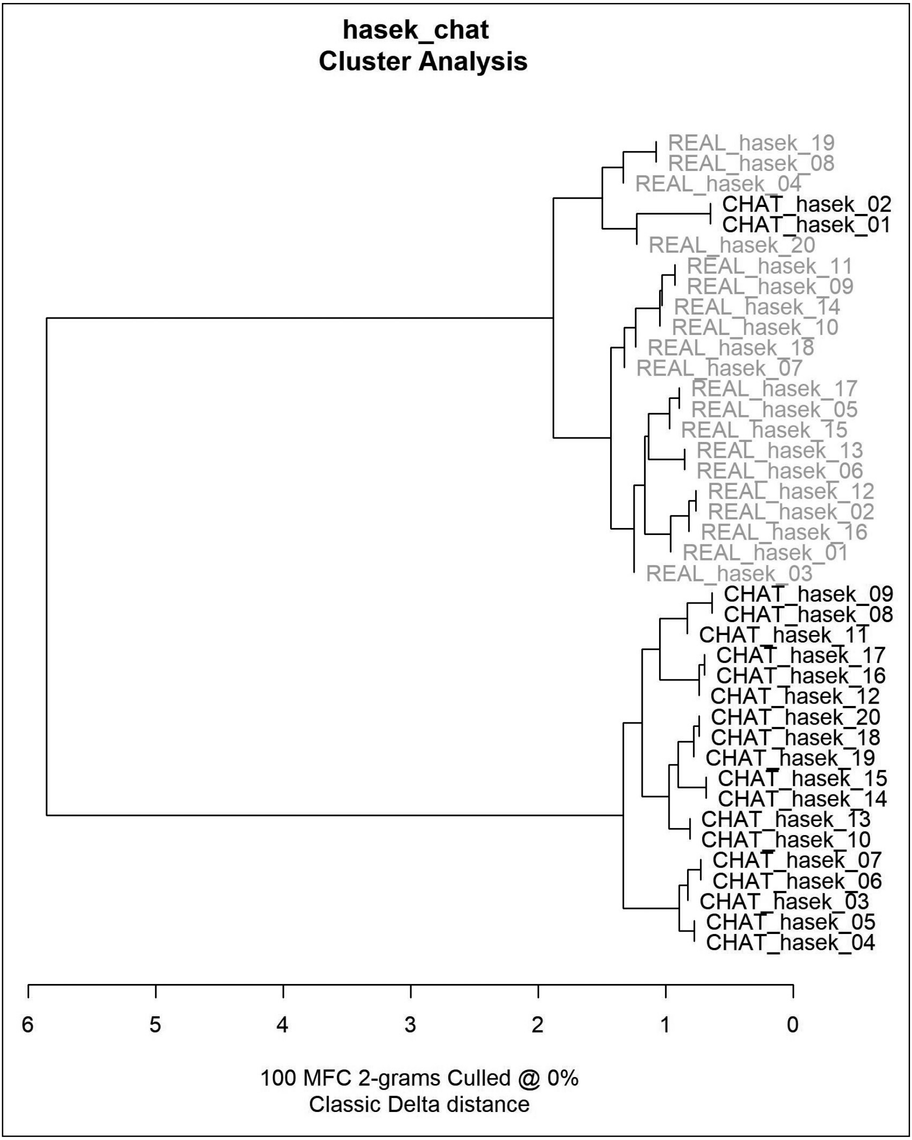

Figure 3. Cluster analysis of the Jaroslav Hašek corpus on the grounds of the 100 most frequent character bigrams and the Classic Delta distance.

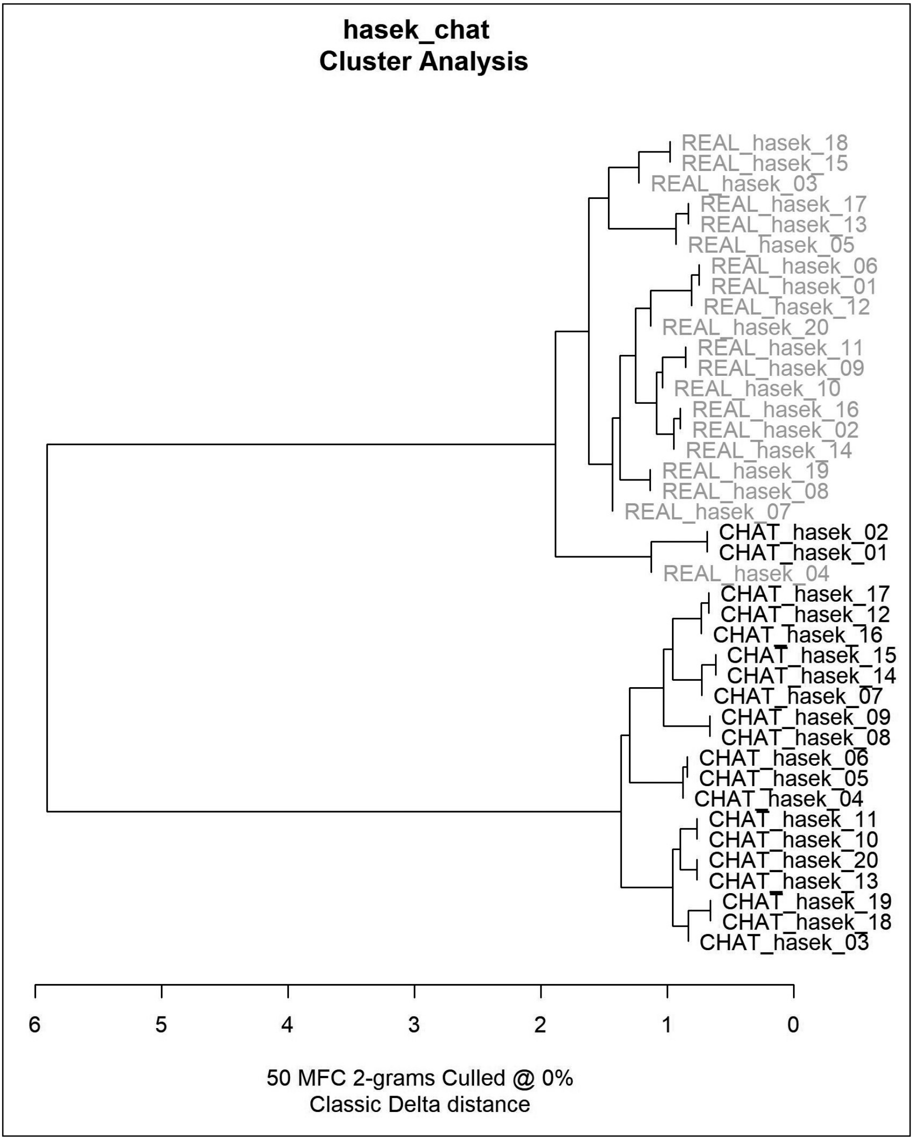

Figure 4. Cluster analysis of the Jaroslav Hašek corpus on the grounds of the 50 most frequent character bigrams and the Classic Delta distance.

4. Conclusions

Given the outcomes of the analysis, it may be preliminarily concluded that the chatbots in their current form are not able to emulate the style of the selected authors successfully. The greatest correspondence was achieved in the case of Karel Čapek, where less than half the indices show statistically significant differences in the values during the text comparisons (14 out of 30). Hašek and Vančura follow with considerably less persuasive numbers (19 out of 30 and 20 out of 30, respectively), and Kafka scores the worst, with 23 out of 30 markers not hit by the GPT. The reason behind these mismatches may be the overall nature of ChatGPT, which tends to produce corporate-sounding texts, or the way it learns the style, since it may use theoretical information rather than the actual training texts. This point has already been discussed above. Concerning the most-frequent-element analyses, the picture is even more one-sided, with the original and chat-produced texts being placed into different clusters in most cases.

The results arrived at by the present investigations have several implications for teaching practice. First and foremost, it appears that the present-day models cannot be used as faithful emulators of the authorial styles that were researched; the teacher should thus always take into account the degree of overlap of the texts produced by models with real ones (see paragraph 1). Employing works by Ernest Hemingway and Mary Shelley, Mikros (2025) arrived at the same conclusion. Second, however, it is possible to utilise chatbots in a confrontational way – comparing GPT-produced texts with originals may sharpen students' stylistic sensibility and lead to recognition and appreciation of an author's hard-to-imitate mastery of style. This approach may be developed further with experimental "triangulation", which would involve contrasting an original text, one produced by a chatbot, and an attempt written by students. It is recommended that such exercises result in editing the texts which were not written by the author so that they would be closer to his/her original style. Chatbots may thus serve to stimulate deep engagement with the authorial style, despite not currently being able to capture it in its complexity. It is thus vital to provide enough training for teachers in order for them to have a good command of the tools offered by generative AI.

The research has several limitations that need to be pointed out. First, the number of samples is rather small, and it is possible that the addition of more texts would provide us with a more comprehensive picture of how the attempt to emulate style works. Second, the prompting may have been misleading for the GPTs, even though there was a strong endeavour to keep the wording the same for all the chatbots; more reminders could have been used for the chatbots to keep in mind that they need to emulate the styles of the given authors. Third, the choice of indices was rather biased towards morphology; more markers linked to syntax, lexis, and text structures may, therefore, reshape the current results and explain the differences (or absence of those) more profoundly.

References

Burrows, J. F. (2002). "Delta": A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, *17*(3), 267–287.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics*, *17*(2), 94–100.

Cvrček, V., Laubeová, Z., Lukeš, D., Poukarová, P., Řehořková, A., & Zasina, A. J. (2020a). *Registry v češtině*. Lidové noviny.

Cvrček, V., Čech, R., & Kubát, M. (2020b). *QuitaUp – nástroj pro kvantitativní stylometrickou analýzu*. Czech National Corpus and University of Ostrava. https://korpus.cz/quitaup/

Dahl, Ö. (Ed.). (2000). *Tense and aspect in the languages of Europe*. Mouton de Gruyter.

Daneš, F. (1954). Příspěvek k poznání jazyka a slohu Haškových „Osudů dobrého vojáka Švejka". *Naše řeč, 37*(3–6), 124–139.

Davidson, D. (2001). *Subjective, intersubjective, objective*. Oxford University Press.

DeepL Translator Team. (2025). *DeepL Translator*. https://www.deepl.com

Eder, M., Rybicki, J., & Kestemont, M. (2016). Stylometry with R: A package for computational text analysis. *The R Journal*, *8*(1), 107–121.

Janda, L. A., Fidler, M., Cvrček, V., & Obukhova, A. (2022). The case for case in Putin's speeches. *Russian Linguistics, 47*(1), 15–40.

Kalantzis, M., & Cope, B. (2025). Literacy in the time of artificial intelligence. *Reading Research Quarterly*, *60*(1), 1–34.

Kosmas, P., Nisiforou, E. A., Kounnapi, E., Sophocleous, S., & Theophanous, G. (2025). Integrating artificial intelligence in literacy lessons for elementary classrooms: A co-design approach. *Educational Technology Research and Development*, *73*(3), 2589–2615.

Kubát, M. (2016). *Kvantitativní analýza žánrů*. Filozofická fakulta Ostravské univerzity.

Kundera, M. (1960). *Umění románu: cesta Vladislava Vančury za velkou epikou*. Československý spisovatel.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*(1), 50–60.

Mikros, G. K. (2025). Beyond the surface: Stylometric analysis of GPT-4o's capacity for literary style imitation. *Digital Scholarship in the Humanities*, *40*(2), 587–600.

Milička, J., Marklová, A., & Cvrček, V. (2025). Benchmark of stylistic variation in LLM-generated texts. *arXiv*, 2509.10179v1.

Místecký, M., & Melka, T. S. (2021). Literary "higher dimensions" quantified: A stylometric study of nine stories. *Glottotheory*, *12*(2), 129–157.

Místecký, M., & Radková, L. (2020). School and gender in numbers: A stylometric insight into the lexis of teenagers' description essays. *Glottometrics*, *49*, 52–65.

Mukařovský, J. (1939). Próza K. Čapka jako lyrická melodie a dialog. *Slovo a slovesnost*, *5*(1), 1–12.

O'Sullivan, J. (2024). Stylometric comparisons of human versus AI-generated creative writing. *Humanities and Social Sciences Communications*, *12*(1), 1708.

OpenAI. (2025). ChatGPT (May 13 version) [Large language model]. https://chat.openai.com/

Piorecký, K., & Husárová, Z. (2018). Tvořivost literatury v éře umělé inteligence. *Česká literatura*, *67*(2), 145–169.

Rebora, S. (2023). GPT-3 vs. Delta: Applying stylometry to large language models. In E. Carbé, G. Lo Piccolo, A. Valenti, & F. Stella (Eds.), *La memoria digitale: Forme del testo e organizzazione della conoscenza. Atti del XII Convegno Annuale AIUCD* (pp. 292–297). Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD).

Mishra, S. (2023). Dissolution of language leads to dissolution of self: Pragmatic analysis of Franz Kafka's Trial. *Indian Journal of Language and Literary Studies, 4*(4), 1–8.

Straková, J., Straka, M., & Hajič, J. (2014). Open-source tools for morphology, lemmatization, POS tagging and named entity recognition. In K. Bontcheva & J. Zhu (Eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 13–18). Association for Computational Linguistics.

Tao, M., Tao, J., & Xu, Q. (2025). A quantitative study on the improvement of students' reading literacy by AI-assisted English reading comprehension training platform. *International Journal of Environmental Sciences*, *11*(20), 1298–1306.

Vondráček, M. (2013). Vlastnosti slova a slovní druhy. In O. Uličný & O. Bláha (Eds.), *Úvahy o české morfologii. Studie k moderní mluvnici češtiny 6* (pp. 17–32). Univerzita Palackého v Olomouci.

Zaitsu, W., & Jin, M. (2023). Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis. *PLOS ONE, 18*(8), e0288453.

Michal Místecký
University of Ostrava, Faculty of Arts, Department of Czech Language;
email: michal.mistecky@osu.cz

MÍSTECKÝ, M. Přeherodovat Heroda? – Modely GPT trénované na autorských dílech a původní tvorba z perspektivy kvantitativní lingvistiky

Cíle: *Studie porovnává texty vytvořené modely GPT trénovanými na dílech významných českých autorů s literárními texty, které tito autoři skutečně napsali. Cílem je zjistit (1) zda mezi oběma typy textů existují rozdíly, a pokud ano, (2) v které jazykové oblasti jsou tyto rozdíly nejvýraznější.*

Metody: *Pro trénink GPT byli zvoleni autoři Karel Čapek, Jaroslav Hašek, Franz Kafka a Vladislav Vančura. Korpus každého autora obsahuje 40 textových vzorků o délce 1 000 slov; 20 je generováno příslušným GPT a 20 převzato z původních děl. Byly provedeny dvě analýzy: první spočívala ve výpočtu 30 morfologických, syntaktických a lexikálních markerů pro každý text, druhá vycházela z analýz nejfrekventovanějších prvků. Výsledky první analýzy byly testovány z hlediska statistické významnosti pomocí Mannova–Whitneyho U testu.*

Výsledky: *Chatboti nedokážou dobře zachytit hovorovost stylu a konverzační interakci a mají tendenci posilovat narativní charakter textu. Nejlepších výsledků je dosaženo u Karla Čapka, nejhorších u Franze Kafky. Stylometrické analýzy téměř vždy dokážou odlišit texty generované umělou inteligencí od textů lidských autorů.*

Závěry: *Texty vytvářené GPT modely trénovanými na konkrétních autorech jsou stále velmi dobře rozlišitelné od textů skutečných spisovatelů. Z hlediska pedagogické praxe však mohou být chatboti využitelní při kritickém srovnávání s původními texty nebo s texty vytvářenými žáky a studenty.*

Klíčová slova: *kvantitativní lingvistika, stylometrie, umělá inteligence, chatbot, ChatGPT, česká literatura*