

Rafał Mazur 
Uniwersytet Jagielloński, Kraków
r.mazur@uj.edu.pl

O POPRAWNOŚCI JĘZYKOWEJ TEKSTÓW GENEROWANYCH PRZEZ SI NA PRZYKŁADZIE CHATUGPT

Słowa kluczowe: błędy językowe, chatbot, przetwarzanie języka naturalnego, sztuczna inteligencja
Keywords: artificial intelligence, chatbot, linguistic mistakes, natural language processing

Możliwości i ograniczenia generatywnej sztucznej inteligencji

ChatGPT to chatbot¹ wykorzystujący mechanizmy generatywnej sztucznej inteligencji (SI). Aplikacja ta została publicznie udostępniona 30 listopada 2022 r. przez firmę OpenAI. Została ona opracowana na podstawie dużego modelu językowego (LLM, od ang. Large Language Model) z rodziny GPT (Generative Pre-trained Transformer). Program jest zdolny do przetwarzania języka naturalnego (NLP, od ang. Natural Language Processing), rozumienia zadanych pytań i generowania odpowiedzi, a także nawiązywania do wcześniejszych wypowiedzi w sposób zbliżony do ludzkiej konwersacji. Wykorzystywany w usłudze model językowy został uprzednio wytrenowany na dużej bazie danych² z użyciem metody nadzorowanego uczenia maszynowego opartego na sieciach neuronowych (jego architektura składa się z tzw. sztucznych neuronów, które strukturą i funkcją przypominają ludzki mózg) (Haque

-
- 1 Program komputerowy, który wchodzi w interakcję z człowiekiem w reakcji na wiadomości tekstowe lub głosowe przesyłane przez użytkownika (Bilińska 2020: 690).
 - 2 Do treningu pierwszego publicznie dostępnego modelu z rodziny GPT (GPT-3) użyto korpusu liczącego 300 miliardów tokenów (Koubaa, Boulila, Ghouti et al. 2023: 3).

2023: 3). Choć model GPT nie jest pierwszą generatywną sztuczną inteligencją (por. badania nad tworzonymi przez komputery obrazami i muzyką sięgają połowy XX w. [por. Mohamadi, Mujtaba, Le et al. 2023: 4]), to wydaje się, że opracowany na jego podstawie chatbot odegra kluczową rolę w rozwoju i społecznej akceptacji technologii związanych z SI. Nie ulega wątpliwości, że ChatGPT szybko stał się powszechnie używanym narzędziem – szacuje się, że pół roku po udostępnieniu usługi stronę openai.com odwiedza miesięcznie ponad 1,8 miliarda osób³. Globalne zainteresowanie chatbotem wiąże się zapewne z dużą liczbą dziedzin, w których użyteczna jest pomoc ChatuGPT⁴, a także z wysokim poziomem zaawansowania tej technologii. SI stworzona przez firmę OpenAI może być wykorzystywana między innymi do: redagowania e-maili i dokumentów aplikacyjnych, pisania scenariuszy filmowych i artykułów badawczych oraz generowania kodu programów komputerowych (Farina, Lavazza 2023). Co więcej, badania udowadniają, że wersja ChatuGPT opracowana na podstawie modelu językowego GPT-4 jest w stanie z powodzeniem rozwiązywać zadania pojawiające się na egzaminach na poziomie szkół wyższych (Newton, Xiromeriti 2023; Gilson, Safranek, Huang et al. 2023). Zdolności komunikacyjne SI opracowanej przez OpenAI ukazują zaś wyniki eksperymentu, w którym wykazano, że chorzy zadający w mediach społecznościowych pytania dotyczące zdrowia oceniają odpowiedzi chatbota jako bardziej empatyczne od porad prawdziwych lekarzy (Ayers, Poliak, Dredze et al. 2023). Co więcej, teksty generowane przez

3 Dane pochodzą ze strony: <https://www.similarweb.com/> (dostęp: 26 II 2024).

4 Nazwa *ChatGPT* funkcjonuje w polszczyźnie od niedawna i nie jest konstrukcją typową dla języka polskiego – stanowi połączenie rzeczownika *chat* oraz nieodmiennego ze względu na formę gramatyczną skrótowca *GPT* (*Generative Pre-trained Transformer*). Warto zwrócić uwagę na to, że wątpliwości użytkowników języka budzi kwestia odmienności wyrazu powstałego z takiego połączenia. Wybór właściwej formy nie jest oczywisty, czego dowodzą sprzeczne rozstrzygnięcia normatywne podawane przez językoznawców. Agata Hącia (2023) stwierdza, że choć rzeczownik *ChatGPT* nie ma wyrazistych cech, które wskazywałyby na jego odmiennność, to przez niektórych użytkowników polszczyzny może być traktowany jako zrost z fleksją wewnętrzną ze względu na gramatyczną odrębność członów składowych. Dopuszczalna byłaby zatem odmiana *ChatGPT* – *ChatuGPT* na zasadzie analogii do np. *Białystok* – *Białegostoku*. Taki wzorzec odmiany zalecają też redaktorzy *Dobrego słownika* (DS) (podkreślają oni jednak, że nazwy tej odmieniać nie trzeba). Innego zdania jest jednak Adam Wolański (2023), który traktuje *ChatGPT* jako wyraz nieodmienny. Badacz argumentuje, że neologizm ten powstał na gruncie angielskim, trudno zatem traktować go jako zrost z fleksją wewnętrzną. Odmienne propozycję przedstawia też Michał Machura (2023). Sugeruje on uznanie omawianej nazwy nie za zrost, lecz zestawienie składające się z rzeczownika i skrótowca (podobnie jak *PKS Opole* lub *numer NIP*), co wiązałoby się z pisownią rozdzielną i możliwością dopuszczenia odmienności pierwszego członu. Do tego rozstrzygnięcia nie przychyliła się jednak ani Wolański, ani redaktorzy *Dobrego słownika* – zapis rozdzielny wydaje się nieuzasadniony jako niezgodny z formą angielską. Wobec różnych stanowisk zajmowanych przez normatywistów sądzę, że najlepszym rozwiązaniem w tym wypadku jest przyjęcie uzusu za podstawę normy. Ponieważ kwerenda językowa przeprowadzona za pomocą wyszukiwarki korpusowej MoncoPL (MPL) potwierdza dużą frekwencję form deklinacyjnych utworzonych zgodnie z wzorcem fleksji wewnętrznej, w tym artykule stosuję taki właśnie zapis.

ChatGPT odznaczają się dużym stopniem kohezji i koherencji, przez co bardzo trudno odróżnić je od tych, które zostały napisane przez człowieka⁵.

Mimo ogromnej bazy danych i dużej mocy obliczeniowej, z których korzysta chatbot, udzielane przez niego odpowiedzi nie są jednak wolne od błędów. ChatGPT niekiedy wykazuje problemy z logicznym rozumowaniem i rozwiązywaniem zadań matematycznych, nie zawsze poprawnie interpretuje humor zawarty w wypowiedziach przesyłanych przez użytkownika, może też popełniać błędy przy generowaniu kodu programistycznego, tworzyć zdania niezgodne z zasadami gramatyki, a nawet podawać nieprawdziwe informacje⁶ (por. Borji 2023). Poprawność odpowiedzi udzielanych przez SI w dużym stopniu zależy od jakości i liczby dostarczonych danych treningowych. Należy zauważyć, że duże modele językowe nie potrafią interpretować przetwarzanych przez siebie informacji tak, jak robią to ludzie. SI próbuje rozwiązać przedstawiony problem w możliwie najlepszy sposób, opierając się na wzorcach utworzonych w procesie głębokiego uczenia, nierzadko powtarzając wartości, poglądy, stereotypy i uprzedzenia zawarte w danych treningowych, bez możliwości weryfikacji ich na podstawie własnych doświadczeń i wiedzy o świecie rzeczywistym. SI, które tak się zachowują (powtarzają przyswojone informacje bez pełnego zrozumienia ich znaczenia i kontekstu), określane są przez badaczy mianem *stochastycznych papug* (Bender, Gebru, McMillan-Major et al. 2021). Do błędu w wypowiedziach generowanych przez SI może prowadzić też brak odpowiedniej liczby danych na temat rozwiązywanego problemu. Na przykład, jak dowodzą badania, efektywność w przetwarzaniu danego języka naturalnego przez ChatGPT jest w dużym stopniu związana z dostępnymi zasobami danych w tym języku – chatbot popełnia najmniej pomyłek, odpowiadając na pytania zadane w języku angielskim (Lai, Ngo, Veyseh et al. 2023).

Można założyć, że wraz z rosnącą popularnością bota stworzonego przez firmę OpenAI, a także innych podobnych aplikacji⁷, w przestrzeni publicznej będzie pojawiać się coraz więcej tekstów tworzonych przy użyciu SI. Jakość takich wypowiedzi może mieć niebagatelny wpływ na kształt uzusu, a przez to na dalszy rozwój normy językowej i całego systemu języka. Zasadne wydaje się zatem zbadanie stopnia poprawności językowej tekstów generowanych przez SI, aby poznać jej możliwości i ograniczenia.

5 Być może w przyszłości posłużą do tego narzędzia z dziedziny stylistyki (por. Szumska 2023: 33) i stylometrii (Zaits, Jin 2023).

6 Przestrzegają przed tym twórcy, którzy na stronie aplikacji informują, że ChatGPT może generować nieprecyzyjne informacje na temat ludzi, miejsc lub faktów (<https://chat.openai.com/>).

7 W trakcie pisania tego artykułu poza aplikacjami stworzonymi na podstawie modelu językowego GPT można było skorzystać także m.in. z chatbota Bard, który został zaprojektowany przez firmę Google.

Metodologia badawcza

Celem badania była ocena umiejętności tworzenia przez ChatGPT wypowiedzi na poziomie egzaminu maturalnego zgodnych z zasadami poprawności językowej. Na potrzeby badania posłużyłem się darmową (można więc założyć, że najczęściej wykorzystywaną przez użytkowników) wersją aplikacji, stworzoną na podstawie modelu językowego GPT-3.5⁸. Teksty były generowane przez SI na podstawie przesłanego przeze mnie zapytania, które zawierało polecenie zaczerpnięte z arkusza z matury z języka polskiego na poziomie podstawowym, rozwiązywanego przez maturzystów w 2023 r.:

Zadanie 18. (0–35)

Wybierz jeden z poniższych tematów i napisz wypracowanie.

- W wypracowaniu rozważ problem podany w temacie. Przedstaw również swoje zdanie i je uzasadnij.
- W rozważaniach przedstaw argumenty, odwołując się do utworów literackich oraz do wybranych kontekstów (np. historycznoliterackiego, literackiego, biograficznego, kulturowego, religijnego, mitologicznego, biblijnego, historycznego, filozoficznego, egzystencjalnego, politycznego, społecznego).
- Jednym z utworów literackich musi być lektura obowiązkowa, wybrana spośród lektur wymienionych na stronach 3 i 4 tego arkusza egzaminacyjnego.
- Twoja praca powinna liczyć co najmniej 300 wyrazów (CKE 2023: 5).

Polecenie ograniczyłem do pierwszego z dwóch dostępnych dla maturzystów tematów:

Temat 1.

Człowiek – istota pełna sprzeczności. W pracy odwołaj się do:

- wybranej lektury obowiązkowej – utworu epickiego albo dramatycznego
- innego utworu literackiego – może to być również utwór poetycki
- wybranych kontekstów (CKE 2023: 5).

Do treści zadania dołączyłem też (stanowiącą część arkusza) listę lektur obowiązkowych w latach 2022/2023 i 2023/2024, aby zapewnić SI kontekst niezbędny do stworzenia wypracowania. Badanie zostało powtórzone w odrębnych konwersacjach, tak by aplikacja nie mogła korzystać z wcześniej wygenerowanych wypowiedzi. W efekcie uzyskałem 25 tekstów o długości od 321 do 449 wyrazów. Teksty te następnie przeanalizowałem pod kątem poprawności językowej. Na potrzeby analizy

8 W trakcie przeprowadzania badania dostępna była też płatna wersja, przygotowana na podstawie bardziej dopracowanego modelu GPT-4.

posłużyłem się typologią Andrzeja Markowskiego (2006: 56–59), który dzieli błędy językowe na:

wewnętrznojęzykowe – systemowe:

- błędy gramatyczne,
- błędy fleksyjne,
- błędy składniowe,
- błędy leksykalne,
- błędy słownikowe (wyrazowe),
- błędy frazeologiczne,
- błędy słowotwórcze,
- błędy fonetyczne;

wewnętrznojęzykowe – stylistyczne;

zewnętrznojęzykowe:

- błędy ortograficzne,
- błędy interpunkcyjne.

W badaniu brałem pod uwagę przede wszystkim błędy w wyraźny sposób naruszające normę językową. Przy klasyfikowaniu analizowanych konstrukcji jako błędów, które w tekście napisanym przez maturzystę mogą razić świadomych użytkowników języka⁹, posługiwałem się przede wszystkim opublikowanymi na stronie OKE¹⁰ materiałami przygotowanymi dla egzaminatorów maturalnych przez Andrzeja Markowskiego (2014), Iwonę Burkacką (2014) oraz Barbarę Pędzich (2014a; 2014b). Założyłem, że wykorzystanie na potrzeby badania kryteriów maturalnych umożliwi porównanie poziomu sprawności językowej tekstów tworzonych przez ludzi (dorosłych użytkowników języka) i generowanych przez SI. Trzeba jednak zaznaczyć, że w analizie nie brałem pod uwagę kwestii pozajęzykowych, które w znaczący sposób nie wpływały na zrozumiałość całego tekstu lub jego fragmentu (np. poprawności merytorycznej generowanych wypowiedzi czy jakości przeprowadzonej argumentacji).

Należy podkreślić, że badanie ma charakter pilotażowy. Wydaje się jednak, iż na jego podstawie można podjąć próbę odpowiedzi na kilka ważnych pytań związanych z poprawnością tekstu generowanego za pomocą ChatuGPT: 1. Jak często w wypowiedziach generowanych przez SI pojawiają się usterki językowe; 2. Jakiego

9 Należy tu zaznaczyć, że uznanie danego elementu językowego za błąd (a więc także i ocena poziomu naruszenia normy) wynika z różnych czynników i ma charakter częściowo subiektywny (por. Markowski 2006: 55; Dąbrowska, Pasięka 2015: 24).

10 <http://www.oke.krakow.pl> (dostęp: 26 II 2024).

typu błędy SI popełnia najczęściej; 3. Czy błędy popełniane przez SI różnią się od pomyłek popełnianych przez ludzi; 4. Czy poszczególne teksty generowane przez SI są porównywalne względem siebie na poziomie poprawności językowej.

Poprawność gramatyczna

Błędy składniowe

Najtrudniejsze dla SI okazało się prawidłowe oddanie składni języka polskiego (76 błędów). Częste pomyłki dotyczyły łączliwości składniowej w zakresie związków zgody (11 błędów) i związków rzędu (18 błędów). Za Hanną Jadacką warto zauważyć, że:

[w]iedza o upodobnianiu się form gramatycznych w obrębie związków zgody uchodzi za elementarną, a jej praktyczne stosowanie [...] nie nastręcza zwykle użytkownikom języka większych kłopotów. Pojawiają się one dopiero wtedy, gdy strukturą przesądzającą o formie orzeczenia ma być konstrukcja pod jakimś względem nietypową, taka np. jak podmiot szeregowy [...] (Jadacka 2006: 139).

Konstrukcje wskazane przez badaczkę jako trudne rzeczywiście nie zawsze były poprawnie interpretowane przez ChatGPT. Na przykład w jednym z wygenerowanych tekstów SI zignorowała regułę, zgodnie z którą orzeczenie przy podmiocie szeregowym zasadniczo powinno mieć formę liczby mnogiej (liczba pojedyncza dopuszczalna jest tylko wtedy, gdy wszystkie części podmiotu są rzeczownikami abstrakcyjnymi mającymi ten sam rodzaj gramatyczny), por.:

1. Jednym z tych utworów, które doskonale ilustrują tę problematykę, **jest dramat** „Antygona” Sofoklesa **oraz poemat** „Konrad Wallenrod” Adama Mickiewicza. (Tekst 11)

Przyczyną błędu składniowego jest odniesienie liczebnika *jeden* do dwóch desygnatów, co uniemożliwiło użycie orzeczenia we właściwej liczbie. Błąd polegający na nieprawidłowym wyborze formy orzeczenia wystąpił także w zdaniu z podmiotem wyrażonym grupą nominalną zawierającą liczebnik, por.:

2. **Wielu** pisarzy i poetów, poprzez swoje utwory literackie, **starali** się zgłębić tę skomplikowaną naturę człowieka. (Tekst 8)

Co istotne, w analizowanych wypowiedziach wykolejenia składniowe pojawiały się także w typowych strukturach podmiotowo-orzeczeniowych, których poprawne zastosowanie nie powinno sprawiać problemu przeciętnemu użytkownikowi języka, por.:

3. Innym utworem, który w pełni ilustruje sprzeczności człowieka, **jest** „Dziady cz. III” Adama Mickiewicza. (Tekst 9)
4. Jego działania są wewnętrznie spreczne – dąży do władzy, ale jednocześnie **gnębi go wyrzuty** sumienia, wynikające z popełnionych zbrodni. (Tekst 16)
5. Natomiast **poezja** Adama Mickiewicza, zwłaszcza utwory „Dziady cz. III” oraz „Konrad Wallenrod”, **przedstawiają** inną perspektywę na sprzeczności człowieka. (Tekst 20)

Ponadto we frazie: „przedstawiają inną perspektywę na sprzeczności człowieka” skontaminowane zostały dwa schematy składniowe. *Perspektywa* może występować w połączeniu *perspektywa* + na CO tylko w znaczeniu ‘przewidywanie lub nadzieja, że coś nastąpi w danej dziedzinie’ (WSJP PAN). W znaczeniu ‘sposób widzenia i oceny czegoś uwzględniający określony punkt widzenia’ dopuszczalne są wyłącznie połączenia *perspektywa* + JAKA i *perspektywa* + KOGO/CZEGO (WSJP PAN).

SI generowała też zdania, w których pojawiały się błędy w związkach między składnikami grupy nominalnej, w której modyfikatory przymiotnikowe powinny być zgodne pod względem przypadku, liczby i rodzaju z formą rzeczownika jako członu konstytutywnego grupy. Tego rodzaju usterki są szczególnie rażące, gdyż, jak zaznacza H. Jadacka, „[b]rak uzgodnienia cech gramatycznych pomiędzy innymi częściami zdania niż podmiot i orzeczenie zdarza się nieporównanie rzadziej” (Jadacka 2006: 147), por.:

6. [...] to właśnie te sprzeczności i walka z nimi definiują nas jako **istoty ludzkie**. (Tekst 3)
7. Adam i Ewa byli stworzeni jako **niewinni istoty**, jednak pod wpływem pokusy złamali Boże przykazanie. (Tekst 4)
8. To zderzenie sprzecznych uczuć i pragnień ukazuje **wewnętrzną rozdwojenie** człowieka. (Tekst 4)

Naruszenia związków rzędu były częstsze niż usterki w zakresie związków zgody. Wynikały one przede wszystkim z nieznamośności wymagań składniowych czasowników i rzeczowników; por.:

9. W „Antygonie” Sofoklesa, konflikt między prawem ludzkim a prawem boskim **stanowi** główną **osią** dramatu. (Tekst 8)
10. Utwory takie jak „Makbet” czy „Dziady cz. III” ukazują tę sprzeczność wewnętrzną, **którą** człowiek może **doświadczać**. (Tekst 13)
11. „Makbet” jest tragedią, która opowiada historię tytułowego bohatera, który popada w **obsesję władzą i zbrodnią**. (Tekst 19)

Niekiedy tego rodzaju błędy wiązały się z brakiem realizacji pełnego schematu składniowego czasownika; por.:

12. Makbet to postać tragiczna, **którą popychają sprzeczne uczucia**, jak chęć zdobycia władzy i poczucie winy. (Tekst 20)

Popychać w znaczeniu ‘skłaniać kogoś do zrobienia czegoś’ wymaga rozbudowania o dopełnienie dalsze: *popychać* + KOGO + do CZEGO | ku CZEMU (WSJP PAN).

Trudne dla SI okazały się również konstrukcje z imiesłowowym równoważnikiem zdania (8 błędów), w których nie zawsze zachowywana była zasada tożsamości podmiotów, por.:

13. Przykładowo, w starożytnej Grecji, o której pisze Jan Parandowski w „Mitologii”, bogowie i ludzie często wchodzili w konflikt ze sobą, **ukazując** wewnętrzne rozdarcie jednostek. (Tekst 15)

Warto zauważyć, że w podsumowaniu pięciu tekstów powtórzył się błąd w konstrukcjach z imiesłowem utworzonym od czasownika *wnioskować*, por.:

14. **Wnioskując**, człowiek jest istotą pełną sprzeczności, a literatura od wieków podkreśla tę kwestię. (Tekst 4)
15. **Wnioskując**, człowiek jako istota pełna sprzeczności jest tematem powszechnie poruszanym w literaturze. (Tekst 6)
16. **Wnioskując**, zarówno „Makbet” Szekspira, jak i „Inny świat” Herlinga-Grudzińskiego, ukazują człowieka jako istotę pełną sprzeczności. (Tekst 7)
17. **Wnioskując**, człowiek jako istota pełna sprzeczności jest nieodłącznym tematem literatury. (Tekst 14)
18. **Wnioskując**, utwory literackie, takie jak „Antygona” Sofoklesa i „Dziady cz. III” Adama Mickiewicza, ukazują człowieka jako istotę pełną sprzeczności. (Tekst 23)

W przytoczonych zdaniach, poza naruszeniem zasady tożsamości podmiotów, razi ograniczenie schematu składniowego równoważnika – czasownik *wnioskować* w znaczeniu ‘rozpatrując fakty lub dane, dochodzić do jakiegoś wniosku’ powinien występować w schemacie: Rz_{osobowy} + *wnioskować* + o CZYM + (SKĄD) lub Rz_{osobowy} + *wnioskować* + że ZDANIE | ZDANIE PYTAJNOZALEŻNE + (SKĄD) (WSJP PAN).

W jednym przypadku wykoślenie w stosowaniu imiesłowowego równoważnika było związane z użyciem w zdaniu nadrzędnym strony biernej (takie usterki I. Burkacka zalicza do błędów nierażących [Burkacka 2014: 10]), por.:

19. W „Księdze Hioba” stawiana jest kwestia niewinności cierpienia, **podważając** wiarę w sprawiedliwość boską, co jest sprzeczne z tradycyjnymi przekonaniem religijnymi. (Tekst 16)

W badanych wypowiedziach powtarzał się także błąd polegający na niepoprawnym użyciu zaimka *gdzie* (7 błędów). Zaimek ten może służyć wyłącznie do wprowadzania zdań podrzędnych okolicznikowych miejsca, por.:

20. Ten wewnętrzny dualizm jest odzwierciedlony w wielu literackich dziełach, w tym w „Makbecie”, **gdzie** postać Lady Makbet również zмага się z przeciwnymi pragnieniami. (Tekst 14)
21. Odnajdujemy go zarówno w utworach epickich, jak i dramatycznych, **gdzie** autorzy starają się ukazać złożoność ludzkiej natury. (Tekst 23)

Stosunkowo liczne w zebranych materiale były też zdania, których zrozumienie utrudniało nadmierne skondensowanie treści. Takie konstrukcje bardzo często sprawiały wrażenie urwanych i niepełnych (18 błędów), por.:

22. Jednak po usłyszeniu przepowiedni czarownic, które wróżą mu przyszłość jako króla, jego natura zaczyna się zmieniać. (Tekst 4)
23. W utworze „Konrad Wallenrod” spotykamy bohatera, który walczy z obcymi najeźdźcami, choć sam należy do tegoż narodu. (Tekst 14)
24. Natomiast w „Dziadach cz. III” Mickiewicz nawiązuje do chrześcijańskiej symboliki, gdzie Konrad jest postacią tragiczną, przypominającą zbawczą ofiarę. (Tekst 21)

W zdaniu 22. elipsa we frazie „Jednak po usłyszeniu [przez niego] przepowiedni czarownic” sprawia, że okolicznik *po usłyszeniu* składniowo odnosi się do rzeczownika *natura*, co powoduje powstanie błędnego znaczenia ‘jego natura usłyszała przepowiednię’. W kolejnym przykładzie właściwego odniesienia nie ma natomiast zaimek *tegoż* – narodowość Konrada nie jest wskazana w cytowanym zdaniu ani we wcześniejszej części tekstu. W zdaniu 24. został pominięty modyfikator, który powinien doprecyzować człon *zbawcza ofiara* (np. *przypominającą zbawczą ofiarę Chrystusa*) – bez niego porównanie wydaje się niekompletne. Dodatkowy błąd składniowy występujący w tym przykładzie wiąże się z niepoprawnym użyciem zaimka *gdzie*.

W badanych tekstach stosunkowo licznie występowały również błędy dotyczące szyku wyrazów (12 błędów). Tego rodzaju wykolejenia dotyczyły przede wszystkim właściwej kolejności modyfikatorów, por.:

25. Jego **postawa romantyczna i buntownicza**¹¹, dążenie do wyzwolenia i odnalezienia swojego miejsca we współczesnym świecie, koliduje z brutalną rzeczywistością. (Tekst 1)
26. Makbet jawi się zatem jako postać pełna **sprzeczności moralnej**. (Tekst 13)
27. „Makbet” Szekspira i „Zbrodnia i kara” Dostojewskiego doskonale ukazują tę sprzeczność, prezentując **bohaterów zmaganie się** z własnymi ambicjami, moralnością i wewnętrznymi konfliktami. (Tekst 5)

W jednym tekście pojawiła się także usterka związana z nienaturalnym szykiem podmiotu i okolicznika, por.:

28. W „Potopie” Henryk Sienkiewicz oraz w balladzie „Romantyczność” Adam Mickiewicz ukazują złożoność natury ludzkiej, której częścią są zarówno cechy pozytywne, jak i negatywne. (Tekst 24)

Błędy fleksyjne

Teksty wygenerowane przez ChatGPT były wolne od błędów natury fleksyjnej poza dwoma wyjątkami, por.:

29. Natomiast „**Kordian**” można interpretować jako obraz polskiego romantyzmu, pełnego sprzeczności i dylematów moralnych w obliczu trudnej rzeczywistości narodowej. (Tekst 1)
30. Jest symbolem walki o wolność, a jednocześnie służy reżimowi, który **nią** ogranicza. (Tekst 11)

W zdaniu 29. usterka wiąże się z nieuzasadnionym brakiem odmiany tytułu omawianego dzieła¹². W kolejnym przykładzie SI niewłaściwie dobrała formę zaimka (zakres użycia form zaimka *niego* / *niej* jest ograniczony wyłącznie do pozycji po przyimku, między innymi: *dla niej*, *bez niego*, *od niej*, *z nim*). W zdaniu razi nadto użycie słowa *reżim* w znaczeniu ‘nieprzyjazny kraj’ (w przytoczonym fragmencie leksem ten stanowi określenie wrogiego Polsce państwa zakonu krzyżackiego).

11 Usterkę można usunąć także przez postawienie przecinka przed członem *romantyczna i buntownicza*, wtedy przymiotniki pełniłyby funkcję określenia predykatywnego.

12 Wydaje się, że SI traktuje tytuły przytoczone w cudzysłowach jako struktury nieodmienne – można w ten sposób wyjaśnić błędną formę orzeczenia w przykładzie nr 3.

Poprawność leksykalna

Błędy słownikowe

W zakresie poprawności leksykalnej większość naruszeń normy językowej można zaklasyfikować do kategorii błędów słownikowych (54 błędy). Najczęściej tego rodzaju usterki wiązały się z naruszeniem łączliwości leksykalnej wyrazów (35 błędów), por.:

31. **Rozważając człowieka** jako istotę pełną sprzeczności, nie sposób nie odnieść się do postaci Makbeta, głównego bohatera tragedii Szekspira. (Tekst 14)

Czasownik *rozważać* (a także utworzony od niego imiesłów) wymaga dopełnienia wyrażonego rzeczownikiem nieosobowym – *rozważać* + CO (np. *rozważać kwestię / problem / pomysł / kandydaturę*; por. WSJP PAN).

32. Makbet, początkowo szlachetny i odważny wojownik, pod wpływem własnych pragnień i podstępного wpływu żony, **popada w mroczne czyny**.

Popadać w znaczeniu ‘przechodzić stopniowo w stan, nazwany przez następujący rzeczownik’ (WSJP PAN) łączy się wyłącznie z rzeczownikami nazywającymi stany psychiczne. W przytoczonym zdaniu razi także bliskie powtórzenie rzeczownika *wpływ*.

33. Przykładem jest biblijna postać Hioba, który jednocześnie **zachowuje** wiarę w Boga i **cierpienie**. (Tekst 17)

Zachowywać w znaczeniu ‘dbać o to, by coś nie zostało zapomniane lub zniszczone’ (WSJP PAN) łączy się tylko z nazwami desygnatów wartościowanych pozytywnie, niewłaściwe jest zatem połączenie **zachowywać cierpienie*.

34. Fragmenty Księgi Rodzaju i Apokalipsy św. Jana przedstawiają dylematy moralne, wybory i konsekwencje, które **dotykają człowieka**. (Tekst 18)

Dotykać w znaczeniu ‘poruszać jakiś temat’ łączy się wyłącznie z rzeczownikami nieosobowymi.

35. Makbet staje się **namiętnie ambitny** i gotów popełnić morderstwo, aby osiągnąć swoje cele. (Tekst 4)

Przysłówek *namiętnie* może łączyć się jedynie z nazwami czynności (np. *namiętnie całować / dyskutować / jeździć konno*).

Wykolejenia leksykalne rzadziej wynikały z użycia wyrazów w niewłaściwym znaczeniu (12 błędów), por.:

36. Na przykład, w Księdze Rodzaju, Adam i Ewa są stworzeni na obraz Boga, ale jednocześnie popełniają grzech pierworodny, który **wynosi** ich poza raj. (Tekst 16)

Czasownik *wynosić* ma znaczenie ‘niosąc coś, wychodzić z tym skądś i w ten sposób usuwać to stamtąd’ (WSJP PAN), niewłaściwe wydaje się użycie go w sensie metaforycznym.

37. Człowiek, istota niezwykle złożona i wielowymiarowa, stanowi **wieczne** źródło zagadnień filozoficznych, psychologicznych i literackich. (Tekst 17)

Wieczny w znaczeniu ‘nieustanny’ odnosi się do rzeczy i zjawisk niepożądanych (np. *wieczna kłótnia, troska*).

38. Innym utworem literackim, który ukazuje sprzeczności człowieka, jest poemat Adama Mickiewicza „Konrad Wallenrod”. Bohaterem utworu jest człowiek, który walczy z **okupantem** [...]. (Tekst 22)

Rzeczownik *okupant* został błędnie wykorzystany w cytowanym zdaniu jako synonim słowa *najeźdźca* ‘ten, kto dokonuje zbrojnego ataku na obszar suwerennego państwa’ (WSJP PAN). Użycie tego określenia (podobnie jak posłużenie się wyrazem *reżim* w przykładzie 31.) ma charakter anachronizmu – nie nadaje się do opisu wydarzeń przedstawionych w cytowanych utworach.

Stosunkowo rzadko dochodziło do pomylenia wyrazów podobnie brzmiących (5 błędów). Wydaje się jednak interesujące, że wszystkie tego rodzaju wykolejenia dotyczyły błędnego zastąpienia słowa *tyran*, por.:

39. Widzimy, jak człowiek, który walczył za króla, staje się krwiożerczym **tyranozaurtem**, gotowym na wszelkie okrucieństwo, by utrzymać się u władzy. (Tekst 1)
40. W „Makbecie” Szekspir prezentuje nam losy tytułowego bohatera, który z jednej strony jest odważnym wojownikiem i lojalnym wasalem, a z drugiej strony poddaje się pokusie władzy, stając się **tyranią**. (Tekst 12)
41. Tytułowy bohater, początkowo waleczny i szlachetny wojownik, pod wpływem ambicji i pragnienia władzy staje się okrutnym **tyranozaurusem**. (Tekst 16)
42. W „Makbecie”, Szekspir przedstawia postać tytułowego bohatera, który jest pełen sprzeczności. Z początku jest on przedstawiony jako szlachetny

- wojownik, jednak pod wpływem zgubnych ambicji i żądzy władzy staje się okrutnym **tyranozaurusem**. (Tekst 20)
43. Tytułowy bohater jest pełen sprzeczności – na początku utworu jawi się jako bohater szlachetny i odważny, lecz z biegiem czasu staje się **tyranią** [...]. (Tekst 22)

W zdaniu 39. dodatkowo została błędnie zmieniona forma wyrazu *tyranozaur*.

Błędy frazeologiczne i słotwórcze

W analizowanym materiale nie zauważyłem rażących naruszeń poprawności frazeologicznej i słotwórczej poza kilkoma wyjątkami. W trzech tekstach pojawił się błąd frazeologiczny polegający na zmianie formy frazemu, por.:

44. Kordian jest postacią pełną niezdecydowania i konfliktów emocjonalnych, co **prowadzi go na drogę samozniszczenia**. (Tekst 3)
45. Antygona reprezentuje wartości boskie, **które stoi na straży**, podczas gdy Kreon kieruje się praktycznymi potrzebami społeczeństwa. (Tekst 11)
46. Ta teza odnajduje potwierdzenie w licznych dziełach literackich, **które biorą pod lupę naturę ludzką** i ukazują jej różne aspekty. (Tekst 20)

W zdaniu 44. błąd polega na połączeniu różnych związków frazeologicznych w jeden (*prowadzić na manowce / sprowadzać na złą drogę i prowadzić do samozniszczenia*). *Prowadzić* w znaczeniu ‘wywoływać określony skutek’ łączy się z przyimkiem DO (WSJP PAN). W kolejnym przykładzie doszło do redukcji formy frazemu: *ktos stoi na straży czegoś*. Co więcej, forma orzeczenia (*stać*) zdania podrzędnego przydawkowego powinna zostać uzgodniona z podmiotem (*wartości*). W zdaniu 46. pojawiła się błędna innowacja frazeologiczna polegająca na wymianie elementu osobowego na nieosobowy: **coś bierze pod lupę kogoś/coś* zamiast *ktos bierze pod lupę kogoś/coś* (por. WSJP PAN).

W jednym tekście użycie frazemu w niewłaściwym kontekście spowodowało uaktywnienie jego znaczenia dosłownego, por.:

47. Wallenrod jawi się jako bohater narodowy, ale jednocześnie jako zdrajca, który działa wbrew własnemu ludowi. Ta sprzeczność w postaci Wallenroda **przedstawia trudności** związane z lojalnością i dążeniem do sprawiedliwości, co jest powszechnym dylematem człowieka. (Tekst 14)

Odczytanie sensu przytoczonego zdania utrudnia udosłownienie wyrażenia *przedstawia trudności*, odnoszącego się zazwyczaj do sytuacji lub czynności wiążącej się z przeszkodami i problemami, takiej, której rozwiązanie wymaga szczególnego wysiłku (np. *Zaopatrzenie w żywność nie przedstawia trudności*).

Poprawność słowotwórcza została naruszona tylko w jednym zdaniu, w którym SI niepoprawnie utworzyła przymiotnik od rzeczownika *władca*, por.:

48. Na przykład w „Iliadzie” Homera, Achilles jest jednocześnie bohaterem i człowiekiem gniewnym, który buntuje się przeciwko **władczyńskiej** woli bogów. (Tekst 17)

Poprawność stylistyczna

Wykolejenia stylistyczne w analizowanych wypowiedziach mają charakter przede wszystkim nieuzasadnionych powtórzeń. ChatGPT w generowanych tekstach stosunkowo często wykorzystywał frazy, które wystąpiły w poleceniu, co prawdopodobnie jest związane ze sposobem przetwarzania tekstu przez SI i ma pomóc chatbotowi w zachowaniu spójności wypowiedzi, por.:

49. Ta **sprzeczność** w jego charakterze ukazuje, jak człowiek może być pełen paradoksów. Podobnie **sprzeczności** w naturze ludzkiej można dostrzec w „Konradzie Wallenrodzie” Mickiewicza. Główny bohater, Konrad Wallenrod, to postać pełna **sprzeczności**. (Tekst 10)

Niekiedy pojawiały się także rażące powtórzenia struktur składniowych, por.:

50. Przyjrzyjmy się zatem tej problematyce, odnosząc się do lektury obowiązkowej, **jaką jest utwór dramatyczny „Antygona” Sofoklesa** oraz innego utworu literackiego, **jakim jest poemat „Pan Tadeusz” Adama Mickiewicza**. (Tekst 8)

Zdecydowanie rzadziej występowały usterki dotyczące nieuzasadnionego użycia archaizmów i kolokwializmów (4 błędy), por. np.:

51. Ponadto, w kontekście Herlinga-Grudzińskiego, opis łagrów radzieckich jest potężnym **przypomnieniem** o destrukcyjnym wpływie totalitaryzmu na naturę człowieka i konsekwencjach, jakie ma dla jednostki i społeczeństwa. (Tekst 7)
52. To wewnętrzne rozdarcie czyni go [Makbeta] istotą pełną sprzeczności – z jednej strony przepełnionym ambicją, a z drugiej – **zahukanym** przez wyrzuty sumienia. (Tekst 14)

Leksem *przypomnienie* użyty w zdaniu 51. w znaczeniu ‘wspomnienie, to, co sobie ktoś przypomina’ jest słowem przestarzałym (por. SJPdOr), razi też połączenie

tego rzeczownika z przymiotnikiem *potężny*. Usterka w zdaniu 52. polega natomiast na użyciu potocznego określenia *zahukany*. Należy ponadto zwrócić uwagę na to, że czasownik *zahukać* ('pozbawić kogoś pewności siebie na skutek upominania i narzucania mu swojego zdania' WSJP PAN) wchodzi zazwyczaj w połączenia z rzeczownikami osobowymi (*zahukany* + przez KOGOŚ). Co więcej, odczytanie właściwego sensu zdania utrudnia brak uzgodnienia rodzaju między rzeczownikiem *sprzecznosci* i jego określeniami **przepełnionym ambicją*, **zahukanym*.

Poprawność interpunkcyjna i ortograficzna

Teksty wygenerowane przez SI zawierały liczne błędy interpunkcyjne (łącznie 48). Większość z nich dotyczyła zbędnego użycia znaku interpunkcyjnego (45 błędów). Częstym błędem, który pojawiał się w zdaniach tworzonych przez ChatGPT, było między innymi niepotrzebne oddzielanie przecinkiem okolicznika otwierającego zdanie, por.:

53. W kontekście tych utworów literackich, warto odwołać się także do kontekstu filozoficznego i egzystencjalnego. (Tekst 3)
54. Przez analizę tych dzieł oraz odwołanie do odpowiednich kontekstów, postaram się ukazać bogactwo wewnętrznych konfliktów człowieka. (Tekst 12)

Do powtarzających się błędów można zaliczyć także stawianie przecinka po zestawieniu *zarówno... jak*, por.:

55. Wnioskując, zarówno „Makbet” Szekspira, jak i „Inny świat” Herlinga-Grudzińskiego, ukazują człowieka jako istotę pełną sprzeczności. (Tekst 7)
56. Podsumowując, zarówno w „Makbecie” Szekspira, jak i w „Konradzie Wallenrodzie” Mickiewicza, autorzy ukazują człowieka jako istotę pełną sprzeczności. (Tekst 10)

Niekiedy usterki interpunkcyjne polegały także na obustronnym wydzieleniu przecinkami wyrazu lub frazy, które nie stanowiły dopowiedzenia, por.:

57. W tragedii Sofoklesa, „Antygona”, konfrontacja pomiędzy tytułową bohaterką a Królem Kreonem ukazuje dwa przeciwstawne podejścia do sprawiedliwości i lojalności. (Tekst 11)

W dwóch tekstach wystąpił błąd polegający na braku dwukropka poprzedzającego wyliczenie, por.:

58. Poprzez odwołanie się do **kontekstów historyczno-literackiego, filozoficznego, religijnego i społecznego**, można lepiej zrozumieć te sprzeczności i głębsze znaczenie, jakie niosą ze sobą te dzieła literackie. (Tekst 19)
59. Wiele utworów literackich analizuje tę złożoną naturę człowieka, przedstawiając go w różnych **kontekstach społecznych, filozoficznych czy egzystencjalnych**. (Tekst 19)

W jednym zdaniu błąd polegał na użyciu niewłaściwego znaku interpunkcyjnego, por.:

60. Przyjrzyjmy się temu zagadnieniu, odwołując się do wybranej lektury obowiązkowej – dramatu Williama Szekspira „Makbet”, oraz innego utworu literackiego, jakim jest poemat Juliusza Słowackiego „Kordian”. (Tekst 1)

Wtrącenie „dramatu Williama Szekspira »Makbet«” powinno być wydzielone z obu stron tym samym znakiem interpunkcyjnym.

ChatGPT nie miał trudności z właściwym stosowaniem reguł ortograficznych. W analizowanym materiale zauważyłem dwie usterki naruszające normę językową w tym zakresie. W obu przypadkach były to błędy literowe: pierwszy to omawiana niepoprawnie wygenerowana forma **tyranozaurt*, drugi polegał na zmianie formy przymiotnika *prorocka*, por.:

61. Jednak, gdy **prorokba** wyrocznia przepowiada mu przyszłość jako króla, Makbet zaczyna być pociągany przez mroczną stronę swojej natury. (Tekst 13)

Dodatkowy błąd polega na nieuprawnionej kondensacji treści we frazie *przyszłość jako króla*.

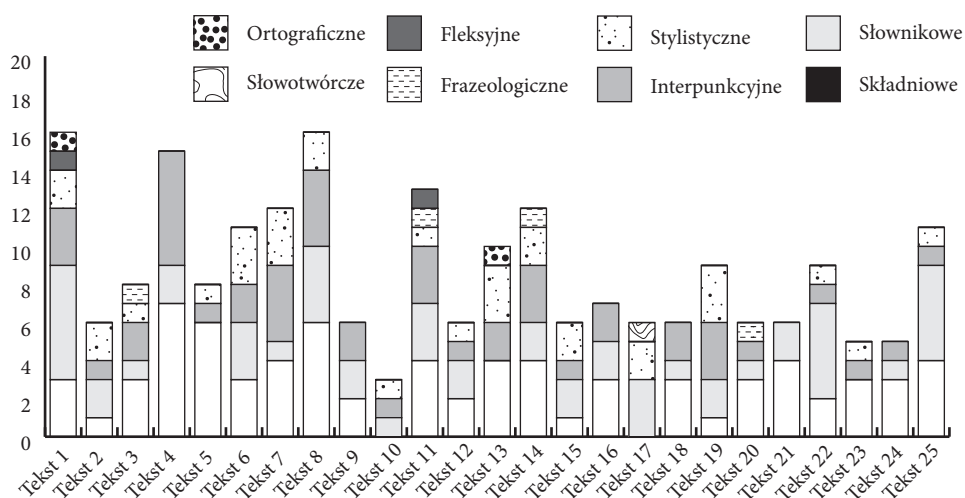
Podsumowanie

Poprawność analizowanych tekstów była różna – liczba usterek językowych w poszczególnych tekstach wygenerowanych przez ChatGPT wahała się w zakresie od 3 do 16 błędów (por. wykres 1).

Najwięcej zastrzeżeń budzi składnia (35% wszystkich błędów). SI nie tylko często posługuje się niewłaściwymi (niekiedy wręcz nienaturalnymi) schematami składniowymi, lecz także nierzadko narusza podstawowe zasady tworzenia zdań, błędnie ustalając formy orzeczeń i podmiotów. Co więcej, nie zawsze potrafi poprawnie uzgodnić formy wyrazów w związkach pobocznych. W badanych tekstach szwankuje także poprawność leksykalna (24% wszystkich błędów). ChatGPT wykazuje tendencję między innymi do tworzenia nietypowych połączeń wyrazowych, czasami

myli również znaczenia słów, których sens powinien być oczywisty dla przeciętnego użytkownika polszczyzny (por. zdania 39–43). W wygenerowanych wypowiedziach licznie występują też naruszenia normy interpunkcyjnej (22% wszystkich błędów), w tym tak rażące, jak oddzielanie przecinkiem podmiotu od orzeczenia. SI dobrze poradziła sobie natomiast ze stworzeniem wypowiedzi neutralnych stylistycznie – do rzadkości należą wyrazy, których nacechowanie stylistyczne odróżnia je od reszty tekstu, na przykład archaizmów czy słów potocznych. Trzeba jednak zaznaczyć, że w badanych tekstach rażą liczne powtórzenia. Być może przyczyną takiego sposobu formułowania wypowiedzi jest to, że chatbota zaprogramowano tak, by często wracał do tematu narzuconego przez użytkownika, co mogłoby pozytywnie wpływać na spójność generowanego przez niego tekstu.

Wykres 1. Liczba błędów w tekstach wygenerowanych przez ChatGPT



Źródło: opracowanie własne.

Nie ulega wątpliwości, że ChatGPT i podobne modele SI są bardzo użytecznymi narzędziami, które mogą znacznie przyspieszyć proces tworzenia tekstu. Wydaje się jednak, że SI, w której treningu wykorzystano technologię dużych modeli językowych, nie może dorównać w sprawności posługiwania się językiem wykształconemu użytkownikowi polszczyzny. Ograniczenia chatbota są widoczne zwłaszcza w tekstach zawierających bardziej złożone i skomplikowane konstrukcje, których ujednolicienie wymaga znajomości szczegółowych zasad poprawnościowych. SI, w przeciwieństwie do człowieka, może w takich wypadkach korzystać wyłącznie z wzorców wyuczonych podczas analizy danych treningowych, wśród których (ze względu na wymóg dostarczenia niezmiernie dużej liczby takich danych) prawdopodobnie nie brak tekstów w mniejszym lub większym stopniu naruszających

normy językowe. Użytkownicy ChatuGPT i innych tego rodzaju narzędzi muszą zatem pamiętać, że tekst wygenerowany przez SI zawsze powinien zostać poddany dokładnej redakcji¹³. Pozostaje mieć nadzieję, że popularność usług wykorzystujących SI nie przyczyni się do rozpowszechniania błędów językowych. Konieczna wydaje się natomiast edukacja społeczeństwa w zakresie odpowiedzialnego posługiwania się tego rodzaju technologiami.

Trzeba w tym miejscu zaznaczyć, że prezentowany tekst należy traktować jako wstęp do dalszych badań – jego celem jest zaznaczenie, moim zdaniem istotnego, problemu badawczego. Ich kontynuacja wymaga powstania zespołu, którego prace mogłyby dotyczyć większej i bardziej zróżnicowanej próby materiałowej – pozwoliłyby na zobjektywizowanie wniosków płynących z analizy. Wartościowe wydaje się między innymi zbadanie poprawności generowanych przez SI tekstów, które należą do różnych gatunków i odmian polszczyzny. Prawdopodobnie możliwe byłoby też wyodrębnienie innowacji swoistych dla SI (błędów, które nie są standardowo popełniane przez użytkowników polszczyzny). Sądzę także, że warto porównywać możliwości wszystkich publicznie dostępnych generatywnych sztucznych inteligencji, ponieważ mogą one mieć w przyszłości znaczny wpływ na rozwój języka.

Źródła

- DS: A. Czesak, Ł. Szałkiewicz, S. Żurowski (red.), *Dobry słownik*, [on-line:] <https://dobrysloownik.pl/>.
 MPL: *MoncoPL*, [on-line:] <http://monco.frazeo.pl/>.
 WSJP PAN: P. Żmigrodzki (red.), *Wielki słownik języka polskiego PAN*, [on-line:] <https://wsjp.pl>.
 ŻMIGRODZKI P., BAŃKO M., BATKO-TOKARZ B. et al. (red.), 2018, *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*, Kraków.

Literatura

- AYERS J.W., POLIAK A., DREDZE M. et al., 2023, *Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum*, „JAMA Internal Medicine” 183 (6), s. 589–596. DOI: 10.1001/jamainternmed.2023.1838.
 BENDER E.M., GEBRU T., McMILLAN-MAJOR A. et al., 2021, *On the dangers of stochastic parrots: Can language models be too big?*, [w:] *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, s. 610–623. DOI: 10.1145/3442188.3445922.

13 Warto podkreślić, że korekta powinna dotyczyć nie tylko formy, lecz także treści. Trzeba również zaznaczyć, że już obecnie w przestrzeni publicznej można zauważyć bardzo dużo tekstów, które prawdopodobnie tworzone są w sposób maszynowy i nie podlegają redakcji i korekcie. Na ich obecność w przestrzeni internetowej jeszcze przed upublicznieniem ChatuGPT zwracał uwagę m.in. Krzysztof Ozga, który analizował artykuły publikowane na portalu Pikio (Ozga 2020: 345).

- BILIŃSKA P.P., 2020, *Chatboty jako narzędzie dystrybucji treści wykorzystywane przez wydawców medialnych*, „Studia Medioznawcze” 21, nr 3 (82), s. 689–707. DOI: 10.33077/uw.24511617.ms.2020.3.186.
- BORJI A., 2023, *A Categorical Archive of ChatGPT Failures*. DOI: 10.21203/rs.3.rs-2895792/v1.
- BURKACKA I., 2014, *Błędy składniowe*, [on-line:] http://www.oke.krakow.pl/inf/filedata/files/Ra%BF%Bice_b%B3%EAdy_j%EAzykowe_2_Sk%B3ad.pdf.
- CKE, 2023, *Egzamin maturalny. Język polski. Poziom podstawowy*, [on-line:] http://cke.gov.pl/images/_EGZAMIN_MATURALNY_OD_2023/Arkusze_egzaminacyjne/2023/Jezyk_polski/poziom_podstawowy/MPOP-P2-100-2305.pdf.
- CZESAK A., SZAŁKIEWICZ Ł., ŻUROWSKI S., 2017, *Dobry słownik jako nowy typ słownika języka polskiego*, „Prace Filologiczne” 71, s. 75–87.
- DĄBROWSKA A., PASIEKA M., 2015, *Błąd językowy – co to takiego? Rozważania o błędzie językowym w glottodydaktyce polonistycznej*, „Acta Universitatis Lodziensis. Kształcenie Polonistyczne Cudzoziemców” 22, s. 21–47. DOI: 10.18778/0860-6587.22.02.
- FARINA M., LAVAZZA A., 2023, *ChatGPT in society: emerging issues*, „Frontiers in Artificial Intelligence” 6. DOI: 10.3389/frai.2023.1130913.
- GILSON A., SAFRANEK C.W., HUANG Th. et al., 2023, *How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment*, „JMIR Medical Education” 9, nr art. e45312. DOI: 10.2196/45312.
- HAQUE M.A., 2023, *A brief analysis of „ChatGPT” – A revolutionary tool designed by OpenAI*, „EAI Endorsed Transactions on AI and Robotics” 1, nr art. e15. DOI: 10.4108/airo.v1i1.2983.
- HĄCIA A., 2023, *ChatGPT*, [on-line:] <https://poradniajezykowa.uw.edu.pl/porady/chatgpt>.
- JADACKA H., 2006, *Kultura języka polskiego. Fleksja, słowotwórstwo, składnia*, Warszawa.
- KOUBAA A., BOULILA W., GHOUTI L. et al., 2023, *Exploring ChatGPT capabilities and limitations: A critical review of the NLP Game Changer*, Preprints.org, nr art. 2023030438. DOI: 10.20944/preprints202303.0438.v1.
- LAI V.D., NGO N.T., VEYSEH A.P.B. et al., 2023, *ChatGPT beyond English: Towards a comprehensive evaluation of large language models in multilingual learning*, arXiv. DOI: 10.48550/arXiv.2304.05613.
- MACHURA M., 2023, *ChatGPT. Dopisek*, [on-line:] <https://sjp.pwn.pl/poradnia/haslo/ChatGPT;22600.html>.
- MARKOWSKI A., 2006, *Kultura języka polskiego. Teoria. Zagadnienia leksykalne*, Warszawa.
- MARKOWSKI A., 2014, *Błędy leksykalne, frazeologiczne i słowotwórcze*, [on-line:] http://www.oke.krakow.pl/inf/filedata/files/Ra%BF%Bice_b%B3%EAdy_j%EAzykowe_1_LeksFrazSlow.pdf.
- MOHAMADI S., MUJTABA G., LE N. et al., 2023, *ChatGPT in the age of generative AI and large language models: A concise survey*, arXiv. DOI: 10.48550/arXiv.2307.04251.
- NEWTON Ph.M., XIROMERITI M., 2023, *ChatGPT performance on MCQ exams in higher education: A pragmatic scoping review*, EdArXiv. DOI: 10.35542/osf.io/sytu3.
- OZGA K., 2020, *Komunikat językowy w pułapce obrazu*, [w:] G. Habrajska (red.), *Teorie i praktyki komunikacji*, Łódź, s. 343–361.
- PĘDZICH B., 2014a, *Błędy stylistyczne*, [on-line:] http://www.oke.krakow.pl/inf/filedata/files/Ra%BF%Bice_b%B3%EAdy_j%EAzykowe_4_Styl.pdf.

- PĘDZICH B., 2014b, *Błędy fleksyjne*, [on-line:] http://www.oke.krakow.pl/inf/filedata/files/Ra%BF%Bice_b%B3%EA%ADy_j%EAzykowe_3_Fleks.pdf.
- PĘZIK P., 2020, *Budowa i zastosowania korpusu monitorującego MoncoPL*, „Forum Lingwistyczne” 7, s. 133–150. DOI: 10.31261/fl.2020.07.11.
- SZUMSKA D., 2023, *Ocalić od zapomnienia: wnikliwość jako kategoria (nie tylko) estetyczna*, „Przegląd Rusycystyczny” 3 (183), s. 31–44.
- WOLAŃSKI A., 2023, *ChatGPT*, [on-line:] <https://sjp.pwn.pl/poradnia/haslo/ChatGPT;22600.html>.
- ZAITSU W., JIN M., 2023, *Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis*, „PLoS ONE” 18 (8). DOI: 10.1371/journal.pone.0288453.

Language Accuracy of Texts Generated by AI: A Case Study of ChatGPT

Abstract

The article aims to assess the linguistic accuracy of texts generated by ChatGPT, based on queries that mirror the instructions found in the 2023 secondary school final exam in Polish language and literature. The analysis focused on identifying and classifying language errors and evaluating their frequency. The results of the analysis revealed a varied level of accuracy in AI-generated texts. Most errors occurred in texts containing more complex constructions, where correct formulation required a comprehensive understanding of grammatical rules. The conclusions drawn from the analysis offer insights into the limitations of AI, which users should be mindful of when editing texts using chatbots.