

Zastosowanie sztucznej inteligencji w edytorstwie naukowym – przykład *Moralistów* Wacława Potockiego

JOANNA HAŁACZKIEWICZ

ORCID: 0000-0001-7143-789X
(Uniwersytet Jagielloński w Krakowie)

KLĘSKA URODZAJU. EDYTORSTWO WOBEC TECHNOLOGICZNEGO POSTĘPU

Gdy pod koniec ubiegłego wieku komputer Deep Blue rozgrywał zwycięskie partie szachowe z arcymistrzem Garrim Kasparowem, świat przyglądał się temu z bez-troskim zaciekawieniem. Oto maszyna stworzona przez człowieka, nakarmiona doświadczeniem wielu wybitnych szachistów, którzy dostarczyli jej gotowych ruchów, zaczęła przewyższać ludzką inteligencję. To wydarzenie ze świata elitarnego sportu i równie niedostępnych laboratoriów Doliny Krzemowej – zupełnie odległe z perspektywy maluczkich – było jaskółką radykalnych zmian technologicznych, które w ostatnich latach skumulowały się i zdaniem wielu zaczęły zagrażać uświęconemu porządkowi¹.

Na przełomie listopada i grudnia 2022 roku jako bezpłatne narzędzie internetowe wystartował ChatGPT. Obywatele globalnej sieci nie mogli przejść obojętnie wobec tego debiutu. Nareszcie pojawił się bot, z którym wszędzie i o każdej porze można porozmawiać na dowolny temat, a przy okazji zaprzęgnąć go do różnorodnych mało ekscytujących prac – innymi słowy: wirtualny towarzysz, osobisty asystent oraz cierpliwy doradca. Dzieło inżynierów z OpenAI odniosło ogromny sukces i pod względem prędkości pozyskiwania nowych użytkowników przewyższyło dotychczasowych liderów w rodzaju TikToka czy Instagrama.

1 Według danych dostarczonych przez firmę GWI dla publikacji *Digital 2024 Global Overview Report* 71% ankietowanych uważa, że rozwój sztucznej inteligencji jest zbyt szybki i budzi niepokój. Data-Reportal, *Digital 2024 Global Overview Report*, <https://datareportal.com/reports/digital-2024-global-overview-report>, s. 31 (wszystkie cytowane w tekście źródła internetowe – stan z 20 września 2024 r.).

Wpływ nowego czatbota okazał się bezprecedensowy², jednak początkowemu entuzjastom zaczął towarzyszyć silny niepokój: a co się stanie, jeżeli wszechstronna i superwydajna sztuczna inteligencja wszystkich nas zastąpi? Już nie tylko szachistów, lecz także lekarzy, prawników, dziennikarzy, artystów, programistów, pracowników biurowych, ekonomistów, tłumaczy, redaktorów czy nauczycieli? Dopóki kosztowne projekty AI pozostawały domeną high-techu, lęk przed utratą pracy czy obniżeniem wynagrodzenia zdawał się nie występować, a przynajmniej nie był on powszechny. W momencie nagłego umasowienia narzędzi wykorzystujących sztuczną inteligencję wielu zaczęło ze strachem myśleć o swojej przyszłości. Sądząc po temacie dziewiętnastej konferencji The European Society for Textual Scholarship w Budapeszcie (*Textual Scholarship, Artificial Intelligence, Corpora and Intelligent Editions*)³, podobna ciekawość podszyta niepewnością towarzyszy również edytorom, choć akurat w tej profesji – zdawałoby się – dość konserwatywnej, świadomość postępu technologicznego oraz udział w nim mają kilkudziesięcioletnią tradycję, zapoczątkowaną jeszcze w latach czterdziestych przez jezuitę Roberto Busę, pomysłodawcę i autora cyfrowego korpusu dzieł świętego Tomasza z Akwinu⁴.

Co czeka edytorstwo naukowe? Jak będziemy pracować nad tekstami dzieł i czy w ogóle będziemy to robić w ten sam sposób co kiedyś? A może sztuczna inteligencja najpierw przeszuka wszystkie dostępne online katalogi bibliotek, dotrze do przekazów zamieszczonych w bibliotekach cyfrowych, automatycznie odczyta ich warstwę tekstową, skolacjonuje zebrany materiał, wychodząc z tych przesłanek, dobierze podstawę edycji, następnie zaś – mając do dyspozycji wiedzę całego internetu, obejmującą także metadyskurs w postaci artykułów o praktyce edytorstwa naukowego – wprowadzi koniektury, skomentuje miejsca wątpliwe, objaśni trudne słowa, pokusi się o erudycyjną przedmowę z wyczerpująco scharakteryzowaną transmisją tekstu? Czy korzystając z już opublikowanych cyfrowych edycji naukowych – przede wszystkim z ich otwartego kodu – wygeneruje własną edycję? Dopóki dworujemy z absurdałnej nieporadności rzekomo potężnego czatbota, który nie umie zliczyć liter „r” w słowie *strawberry*⁵, ta wizja wydaje się odległa, z drugiej

2 Na wykresie Google Trends, rejestrującym zainteresowanie internautów różnymi tematami, widać wyraźnie, jak hasło „sztuczna inteligencja” zanotowało skokowy wzrost popularności pod koniec 2022 r. Według innego raportu w 2023 r. „ChatGPT” był najczęściej czytany hasłem w Wikipedii (ibidem, s. 108). Warto nadmienić, że OpenAI nie jest jedynym dostawcą sztucznej inteligencji – Chiny rozwijają własny projekt pod nazwą Wu Dao, swoje miejsce starają się znaleźć mniejsze start-upy, takie jak Cohere, nad technologią pracują także dotychczasowi giganci IT pokroju Microsoftu (Copilot), Google’a (Gemini) czy Facebooka (Meta AI), obok nich dostęp do nowej technologii demokratyzują organizacje non profit, jak chociażby EleutherAI czy ostatnio Mozilla AI.

3 *Program of the ESTS Conference 2024*, <https://elte-dh.hu/ests-2024-program/>.

4 Corpus Thomisticum, <https://www.corpusthomicum.org/>.

5 A. Silberling, *Why AI Can't Spell 'Strawberry'*, <https://techcrunch.com/2024/08/27/why-ai-cant-spell-strawberry/>.

strony inżynierowie sztucznej inteligencji już poszukują rozwiązań, aby stworzyć nowe modele zdolne do prowadzenia wieloetapowych badań⁶. Dowodem na to, że „sztuczne” jest coraz bardziej „ludzkie” i realistyczne, niech będzie fakt, że przeciętni internauci wątpią w swoje zdolności poznawcze, gdy w mediach społecznościowych przychodzi im odróżnić bota od człowieka, deepfake wygenerowany przez Sorę⁷ od prawdziwego nagrania wideo.

Z programu konferencji ESTS można wywnioskować, że środowisko edytorów obecnie wymienia się pomysłami na to, jak wykorzystać możliwości AI do tworzenia edycji naukowych. Już nie cyfrowe edycje naukowe (*digital scholarly editions*), lecz inteligentne edycje (*intelligent editions, AI editions*) lub chociaż wspomagane przez sztuczną inteligencję (*AI-driven*) stają się najsilniejszą pobudką dla filologicznej wyobraźni⁸. Wydaje się, że to podejście nie odbiega od ogólnego trendu w innych profesjach „zagrożonych” rywalizacją z inteligentnymi komputerami: zamiast obstawać przy z góry przegranej pozycji, lepiej zacząć używać AI do swoich celów, stać się jej operatorem. Zainteresowanie sztuczną inteligencją wśród humanistów wzmacnia się tym bardziej, że wspiera je silny nurt badań ilościowych literatury spod znaku *distant reading*, który teksty literackie traktuje jak wielkie zbiory danych do przetworzenia (*big data*) – podobnie przecież działa generatywna sztuczna inteligencja: swoją „mądrość” czerpie z ogromnych zasobów danych. Coraz głośniejsze jest też o edytorstwie źródeł cyfrowych (*born-digital*), co nie powinno zaskakiwać, gdy weźmiemy pod uwagę, że od ponad półwiecza ludzkość na szeroka skalę tworzy różne dokumenty elektroniczne, a od ponad trzydziestu lat istnieje usługa znana jako globalna sieć (World Wide Web). W sieci i na dyskach komputerów znajdują się niezliczone utwory cyfrowe – wśród nich materiały z pewnością ważne dla badaczy literatury, które domagają się właściwego potraktowania, jednak dotarcie do nich może okazać się niemożliwe bez pomocy asystenta sztucznej inteligencji.

W tej sytuacji na uniwersytetach oraz w instytutach badawczych panuje klęska urodzaju. W wielu miejscach niezależnie od siebie powstają narzędzia i platformy, często niedoskonałe, będące „w trakcie budowy”, „aktualnie przenoszone na serwery chmurowe”, „tylko częściowo dostępne”, za którymi nie sposób nadążyć i któ-

6 A. Tong, K. Paul, *Exclusive: OpenAI Working on New Reasoning Technology under Code Name 'Strawberry'*, <https://www.reuters.com/technology/artificial-intelligence/openai-working-new-reasoning-technology-under-code-name-strawberry-2024-07-12>.

7 Sora to kolejne dzieło inżynierii OpenAI, zdolne generować realistyczne filmy bardzo trudne do odróżnienia od materiałów nagranych tradycyjnie.

8 Podczas konferencji DSE Communities w IBL PAN w Warszawie (25–27 września 2024 r.) Michael Pidd przedstawił założenia projektu C21 Editions, w ramach którego przy szerokim zastosowaniu sztucznej inteligencji powstaje cyfrowa edycja naukowa *Opowieści kanterberyjskich*. Por. C21 Editions. Scholarly Editing and Publishing in the Digital Age, <https://www.c21editions.org/>.

rych na ogół nie da się bez trudu wdrożyć do własnych przedsięwzięć edytorskich, gdyż nie są one dostatecznie wszechstronne lub przyjazne w obsłudze. Pewne wyobrażenie o tym, jak liczne i różnorodne są to zasoby, daje katalog Social Sciences and Humanities Open Marketplace⁹.

Mając do dyspozycji tak wiele narzędzi, edytorzy usiłują na nowo zorganizować swój warsztat i zbudować standardy pracy w środowisku cyfrowym. Miałyby to polegać na starannym wyselekcjonowaniu już istniejących serwisów, aplikacji lub nawet użytecznych skryptów – na przykład do automatycznego transkrybowania, kolacjonowania czy anotowania – po które można później dowolnie sięgać w zależności od potrzeby. Takie próby owocują studiami przypadków, w których autorzy opisują wykorzystanie konkretnych narzędzi do tworzenia edycji.

Wobec tak dużej dynamiki istotne są także rozważania nad długim trwaniem cyfrowych projektów edytorskich (*future-proof editing*), słusznie bowiem dostrzega się zagrożenia płynące z przeładowania edycji oryginalnymi, kosztownymi rozwiązaniami skrojonymi na miarę¹⁰, które po kilku latach przestają się poprawnie wyświetlać w oknach przeglądark. Zdaniem niektórych odpowiedzią na te wyzwania jest tak zwane minimalne przetwarzanie, polegające na korzystaniu z jak najbardziej standardowych języków programowania i kodowania, możliwie niezależnych od zmieniających się standardów internetu¹¹; powstają też repozytoria danych badawczych, w których można zdeponować kod źródłowy edycji.

Brak jednej wygodnej metody edytorstwa cyfrowego, będący przyczyną zniechęcenia wielu badaczy, bywa porównywany z początkami drukarstwa – wówczas także nie było katalogu dobrych praktyk tworzenia inkunabułów, a nawet reguł zapisu języków wernakularnych, jednak mimo to wynalazek Gutenberga stopniowo zyskiwał akceptację, zmieniał stosunki społeczne oraz kulturę intelektualną kolejnych pokoleń czytelników¹². Jak pisał Peter L. Shillingsburg:

» Łatwo pogubić się lub zniechęcić w dziedzinie, jaką tworzą teksty elektroniczne. Każda nowinka w tym świecie szybko staje się przytłaczającym obliczem kolejnych nowości. Kusi, by przeczekać to zamieszanie i dołączyć dopiero do tych, którzy osiągnęli sukces – tak jak

9 SSH Open Marketplace, <https://marketplace.sshopencloud.eu/>.

10 Por. E. Pierazzo, *What Future for Digital Scholarly Editions? From Haute Couture to Prêt-à-Porter*, „International Journal of Digital Humanities” 2019, nr 13, s. 209–220.

11 R. Viglianti, G. del Rio Riande, *Against Infrastructure. Global Approaches to Digital Scholarly Editing*, „C21 Digital Editions” 2023; eidem, N. Hernández, R. De León, *Otwarte, minimalne i promujące równe szanse. Jak uczy się tworzenia naukowych edycji cyfrowych na Globalnej Północy i Globalnym Południu*, tłum. J. Wełniak, „Sztuka Edycji” 2023, nr 1, s. 71–83.

12 H. Hollender, *Czy świat czeka przyszłość średniowiecza?*, w: E.L. Eisenstein, *Rewolucja Gutenberga*, tłum. H. Hollender, Warszawa 2004, passim.

oszuści dołączający w maratonach do przebiegających ostatnie mile. Meta jednak zdaje się oddalać, tak jak horyzont¹³.

Wydaje się, że pojawienie się nowego czynnika w postaci sztucznej inteligencji znacząco zmienia dotychczasowe reguły gry, i właśnie dlatego pokusa przeczekania jest czymś, czemu nie należy zbytnio ulegać.

AI W PRAKTYCE, CZYLI JAK NAUCZYĆ KOMPUTER CZYTAĆ BAROKOWE RĘKOPISY

Utworzony na początku 2023 roku na Uniwersytecie Jagiellońskim zespół Laboratorium Edytorstwa Cyfrowego (LabEdyt)¹⁴ stawia przed sobą kilka zadań: eksperymentowanie z dostępnymi narzędziami w celu zorganizowania systemów przepływu pracy, wymianę doświadczeń w miarę opracowywania naukowych edycji cyfrowych, stałe monitorowanie nowości technologicznych, prowadzenie zajęć dydaktycznych dla studentów, a ponadto wspieranie naukowców w realizacji przedsięwzięć cyfrowych. Obecnie LabEdyt pracuje nad kilkoma projektami pilotażowymi¹⁵, spośród których jeden – poświęcony *Moralium* Wacława Potockiego – bada możliwości wykorzystania uczenia maszynowego do automatycznego tworzenia transliteracji, transkrypcji oraz tagowania znacznikami semantycznymi XML.

XVII stulecie w Rzeczpospolitej Obojga Narodów zdaniem badaczy tego okresu było „wiekiem rękopisów”. Jak pisał Radosław Grześkowiak: „najciekawsze dzieła epoki były powierzane manuskryptom i powielane w obiegu nieformalnym”. Dotyczyło to „twórczości nie tylko takich tuzów jak Jan Andrzej Morsztyn, Wacław Potocki czy Stanisław Herakliusz Lubomirski, ale też ważnych dla dziejów naszej literatury statystów drugiego planu jak choćby Daniel Naborowski czy Hieronim i Zbigniew Morsztynowie”¹⁶. Wacław Potocki (1621–1696), niezwykle płodny literacko szlachcic z Łużnej, porównywany pod tym względem z Józefem Ignacym Kraszewskim, praktycznie wszystkie swoje utwory pozostawił w rękopisach – w tym

13 P.L. Shillingsburg, *Od Gutenberga do Google'a. Elektroniczne reprezentacje tekstów literackich*, tłum. P. Bem, Warszawa 2020, s. 24.

14 Zespół koordynowany przez Magdalenę Komorowską działa w ramach projektu flagowego Digital Humanities Lab, zob. <https://dhlab.id.uj.edu.pl/labedyt>; <https://labedyt.dhlab.uj.edu.pl>.

15 Są to: rewitalizacja Biblioteki Literatury Staropolskiej i Nowołacińskiej „Neolatina” (<https://neolatina.dhlab.uj.edu.pl>), cyfrowa edycja naukowa *Moralioń* Wacława Potockiego oraz edycja korespondencji Szymona Laksa z Krystyną i Czesławem Bednarzykami, właścicielami Oficyny Poetów i Malarzy w Londynie.

16 R. Grześkowiak, *Stary druk jako podstawa edycji krytycznej. Preliminaria*, w: *Jak wydawać teksty dawne*, red. K. Borowiec et al., Poznań 2017, s. 11.

najbardziej znany poemat epicki *Transakcja wojny chocimskiej*. U schyłku jego życia drukiem ukazało się tylko jedno większe dzieło – *Poczet herbów szlachty Korony Polskiej i Wielkiego Księstwa Litewskiego* (wydane w 1696 roku w krakowskiej drukarni Mikołaja Aleksandra Schedla). Pozostałe teksty, w tym *Moralia abo rzeczy do obyczajów nauk i przestróg w każdym stanie żywota ludzkiego z łacińskich i z polskich przypowieści ojczystym krótko napisane wierszem*, czekały na wydawcę do następnych stuleci. Bogatą twórczością staropolskiego pisarza zainteresowano się dopiero na przełomie XIX i XX wieku, zaś największy wkład w jej popularyzację wniósł wówczas Aleksander Brückner¹⁷.

Potocki, mając do dyspozycji edycję *Adagiów* Erazma z Rotterdamu wydrukowaną w 1551 roku u Frobenia w Bazylei¹⁸, od około 1688 roku aż do śmierci pracował nad swoim najobszerniejszym dziełem – *Moraliami*, czyli zbiorem krótkich wierszowanych utworów skomponowanych wokół wybranych i skomentowanych przez Erazma starożytnych maksym. Sporządzony ręką Potockiego czystopis *Moraliońw*, przechowywany w Bibliotece Narodowej¹⁹, rozrósł się do imponujących rozmiarów – 712 kart, czyli 1424 stron, z kolei egzemplarz *Adagiów*, którym pisarz się wspomagał, czego dowodzą jego odręczne notatki na marginesach, został odnaleziony wśród dubletów Biblioteki Jagiellońskiej przekazanych następnie do Towarzystwa Naukowego Warszawskiego (dziś Biblioteka IBL PAN)²⁰.

Dzieło do tej pory doczekało się jednego pełnego wydania – zostało opracowane przez Tadeusza Grabowskiego oraz Jana Łosia i opublikowane w trzech tomach w latach 1915–1918 w serii Biblioteki Pisarzy Polskich. Edycja ta, już ponadstuletnia, a zatem dla współczesnego czytelnika zabytkowa, od początku była krytykowana jako niedokładna i zbyt oszczędna w objaśnieniach. Wiele niedociągnięć wytknął jej Aleksander Brückner²¹, listę omyłek uzupełnił znawca twórczości sarmackiego poety Leszek Kukulski²². Brückner, jakby miał dar jasnowidzenia, pesymistycznie wieszczył, że dzieło Potockiego nieprędko ukaże się w poprawionym opracowaniu: „więc zbytek oszczędności osiągnięto kosztem zrozumiałości, czego

17 Por. Potocki Wacław (1621–1696), w: *Bibliografia literatury polskiej. „Nowy Korbut”*, red. K. Budzyk, t. 3: *Piśmiennictwo staropolskie. Hasła osobowe N–Z*, oprac. R. Pollak et al., Warszawa 1965, s. 119–130.

18 L. Kukulski, *Prolegomena filologiczne do twórczości Wacława Potockiego*, Wrocław 1962, s. 7.

19 W. Potocki, *Moralia*, rękopis, ok. 1688–1696, Biblioteka Narodowa, rps 3049 III, Polona.pl.

20 E. Roterodamus, *Adagiorum Chilides Des. Erasmi Roterodami Quatvor Cvm Dimidia Ex Postrema Avtoris Recognitione: In hac aeditione, prioribus tribus Indicibus subiunctus est quartus nouus [...]*, <https://rcin.org.pl/dlibra/publication/74578/edition/66983/content>. Por. L. Kukulski, *Prolegomena filologiczne...*, s. 7, przyp. 10.

21 A. Brückner, *Wacława Potockiego „Moralia” (1688)*, wyd. Tadeusz Grabowski, Jan Łoś, [Kraków] 1915–1918 [recenzja], „Pamiętnik Literacki” 1920, t. 17/18, nr 1–4, s. 159–164.

22 Zob. np. L. Kukulski, *Prolegomena filologiczne...*, s. 14, przyp. 28.

bardzo żałować należy, bo nie doczekają się chyba *Moralia* lepszego, staranniejszego wydania²³ – pisał i nie mylił się. Fragmenty tego zbioru ukazywały się już później tylko w wyborach, między innymi w trzecim tomie obszernej edycji pism Potockiego²⁴.

Powtórne sporządzenie edycji *Moralioów* to zadanie niewdzięczne z wielu powodów. Przede wszystkim nie jest to „ogród nieplewiony”, by sparafrazować tytuł innego zbioru Potockiego. Z pewnością większa pokusa tkwi w przecieraniu szlaków i zajmowaniu się tekstami nieopublikowanymi. *Moralia*, mając już „jakąś” edycję, przegrywają rywalizację z dziełami wciąż oczekującymi na wydanie. Poza tym, chcąc je opracować, należałoby sięgnąć do rękopisu, za czym edytorzy, jak zdiagnozował Radosław Grześkowiak, najwyraźniej nie przepadają²⁵. Jeśli dodać do tego nieprzeciętne rozmiary tekstu (od razu warto zadać sobie pytanie: jak przekonać komisje grantowe do sfinansowania druku?), konieczność badania powiązań z *Adagiami* oraz fakt, że gros pracy edytorskiej wykonał przed śmiercią Leszek Kukulski (jemu też udało się poczynić najciekawsze odkrycia), to łatwo dojść do pragmatycznej konstatacji, że w czasach naukowego pośpiechu trudno byłoby poświęcić tyle czasu na studiowanie jednego dzieła²⁶.

Nadzieją dla *Moralioów* stają się jednak narzędzia edytorstwa cyfrowego. Dzięki nim możliwe jest stworzenie edycji szybciej i dokładniej, przy niższych kosztach. Takie e-wydanie dodatkowo poszerza możliwości druku, gdyż można w nim zawrzeć różne wersje tekstu (transliterację, transkrypcję modernizującą, tekst z komentarzem edytorskim, faksymile rękopisu, być może także podobizny Frobenowej edycji *Adagiów*), stworzyć interaktywny „klucz do *Moralioów*”²⁷, udostępnić tematyczne indeksy, moduł zaawansowanego przeszukiwania czy choćby listę frekwencyjną – wszystko to bez liczenia się z arkuszami drukarskimi. Odpowiednie oznakowanie tekstu tagami oraz zintegrowanie go z już istniejącymi bazami (na przykład WikiData) dodatkowo włącza go do sieci Linked Open Data – czyli otwartych, powiązanych ze sobą zasobów danych udostępnianych w internecie. Utworzenie takich połączeń daje możliwość prowadzenia dalszych badań, złasz-

23 A. Brückner, *Wacław Potocki „Moralia”...*, s. 161.

24 W. Potocki, *Dzieła*, t. 3: *Moralia i inne utwory z lat 1688–1696*, oprac. L. Kukulski, Warszawa 1987. Bardziej szczegółowy przegląd najnowszych dokonań edytorskich w dziedzinie *potocianów* zob. J. Gruchała, *Wacław Potocki – problem edytorski*, w: *Wacław Potocki. W 400-lecie urodzin poety*, red. K. Koehler, D. Chempererek, Warszawa 2023, s. 15–33.

25 R. Grześkowiak, *Stary druk...*, s. 12.

26 Wprost napisał o tym S. Grzeszczuk: „Potocki ma beznadziejną przewagę nad indywidualnym badaczem, choćby nie wiem jak pracowitym i oddanym sprawie. Potockiego może pokonać tylko zespół [...]” – *O potrzebie i programie badań nad twórczością Wacława Potockiego*, w: *Wśród zagadnień polskiej literatury barokowej*, cz. 2: *Motywy – inspiracje – recepcja*, red. Z.J. Nowak, Katowice 1980, s. 16.

27 Por. L. Kukulski, *Klucz do „Moralioów”*, w: idem, *Prolegomena filologiczne...*, s. 64–68.

cza przy użyciu narzędzi AI. Chatboty korzystające z dużych modeli językowych, jak choćby polski Bielik AI, również są w stanie odczytywać edycję oraz wspomagać jej użytkowników. Ponieważ otagowane fragmenty tekstu mają nadwyżkę semantyczną w postaci metadanych, mogą być przetwarzane przez komputer w sposób bardziej zaawansowany niż w przypadku czystego tekstu. Z tak podanej edycji *Moralioń* użytek mogliby czerpać chociażby twórcy korpusów językowych takich jak KorBa²⁸, gdzie wprawdzie znajdują się wyimki z dzieł Potockiego, lecz są one podane w formie uwspółcześnionej, jak stwierdził Janusz Gruchała: „właściwej raczej edycjom popularnonaukowym”²⁹. Wszelako, aby móc zacząć myśleć o dobrodziejstwach edycji cyfrowych, najpierw należy pozyskać do nich tekst.

Transliteracja, zwana też „transkrypcją dyplomatyczną”, stoi u źródeł każdej edycji tekstów „urodzonych” przed erą cyfrową – a i pewnie w przyszłości będzie dotyczyć wielu utworów wprawdzie napisanych na komputerze, jednak w formie cyfrowej niezachowanych. O jej niepodważalnym znaczeniu pisało wielu. Peter Shillingsburg określił transliterację formą „reinkarnacji” – powtórnego wcielenia pewnej niematerialnej idei, czyli tekstu. „Reinkarnacja” jest więc przyjęciem przez tekst nowego „ciała” (jakkolwiek zastanawiająco brzmi to w kontekście przestrzeni cyfrowej)³⁰. W edytorstwie zgodnie z platońską koncepcją to ucieleśnienie doskonałej idei staje się jednocześnie jej skażeniem, ponieważ materia zawsze pozostaje niedoskonała. Na etapie „reinkarnacji” nie sposób ustrzec się błędów – i popełnia je każdy: kopista, redaktor, korektor, zecer i metrampaż oraz ich współczesny odpowiednik – grafik DTP, drukarz, introligator, a nawet autor, dopuszczając się niezamierzonych potknięć pióra.

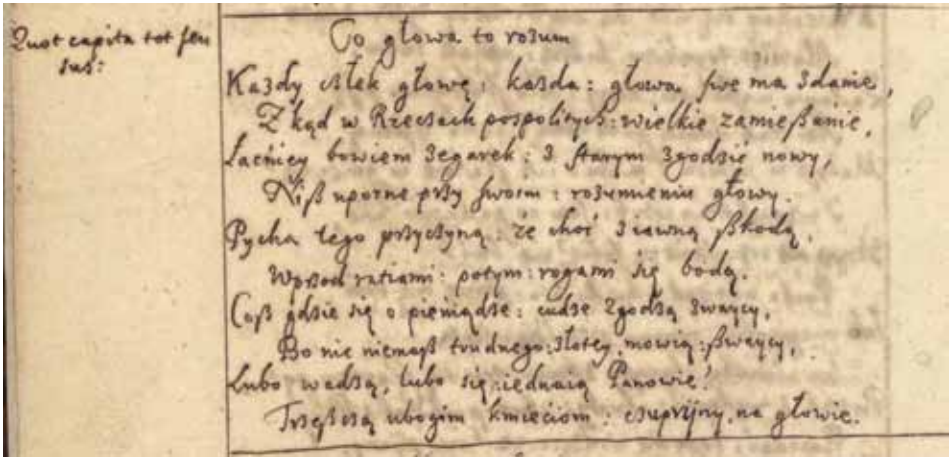
W przypadku *Moralioń* nietrudno o błąd przy przepisywaniu. Ogrom dość jednostajnego materiału sprzyja pomyłkom. Jest to tym prawdopodobniejsze, im więcej osób zajmuje się kopiowaniem jednego dzieła – istnieje ryzyko, że nie wszyscy mimo najszczerzych chęci będą w stanie przestrzegać przyjętych ustaleń. Dodatkowo *Moralia* to tekst z perspektywy dzisiejszego czytelnika zawiły, naszpikowany archaizmami, przez co miejscami wręcz niezrozumiały – ogromna w tym zasługa barokowej poetyki. Taką opinię o dziele miał już Aleksander Brückner sto lat temu³¹, cóż więc ma powiedzieć użytkownik polszczyzny XXI wieku? Brak zrozumienia tekstu może prowadzić do omyłkowych, pochopnych odczytań, a forma przekazu – rękopis – podnosi trudność zadania, choć trzeba podkreślić, że dukt pisma Potockiego jest czytelny.

28 Elektroniczny korpus tekstów polskich XVII i XVIII w. „KorBa”, <https://korba.edu.pl>.

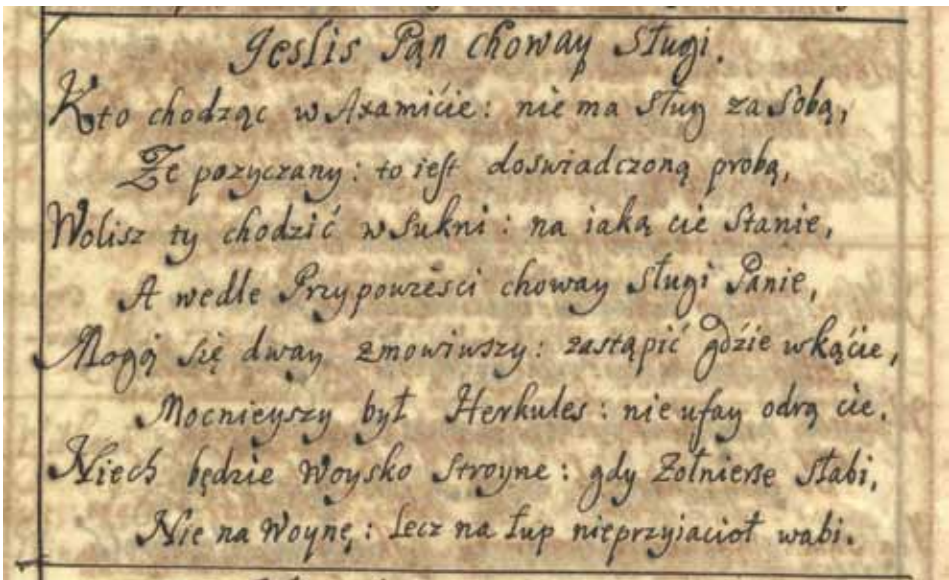
29 J.S. Gruchała, *Wacław Potocki...*, s. 19.

30 P.L. Shillingsburg, *Od Gutenberga...*, s. 27.

31 A. Brückner, *Wacława Potockiego „Moralia”...*, s. 159–161.



Il. 1. Próbką charakteru pisma z początkowych kart rękopisu *Moralioń*. Źródło: Polona (skan 4v–5r).



Il. 2. Próbką charakteru pisma z końcowych kart rękopisu *Moralioń* – fragmenty spisane ręką „chłopca do przepisywania” pochodzą z *Ogródu fraszek*³². Źródło: Polona (skan 641v–642r).

Mając na uwadze te zastrzeżenia, zespół Laboratorium Edytorstwa Cyfrowego podjął decyzję o wykonaniu transliteracji za pomocą narzędzia do automatycznego rozpoznawania tekstów rękopiśmiennych – aplikacji Transkribus. Program ten (choć w aktualnym stanie rozwoju należałoby go chyba nazwać narzędziem SaaS –

32 Por. J. Łoś, *Wstęp*, w: *Wacława Potockiego „Moralia” 1688*, t. 3, wyd. T. Grabowski, J. Łoś, Kraków 1918, s. XXVI.

Software as a Service – gdyż tak zwany klient desktopowy został już wycofany) wykorzystuje spreparowane przez użytkowników próbki obrazów i powiązanych z nimi transkrypcji do trenowania wyspecjalizowanych modeli odczytywania tekstów. Sam „trening” polega na zastosowaniu podstawowej funkcji sztucznej inteligencji, jaką jest uczenie maszynowe na dostarczonym zbiorze uczącym. Wytrenowany model można zapisać w prywatnej bibliotece przyporządkowanej do konta lub udostępnić go Transkribusowej społeczności.

W aplikacji dostępnych jest wiele gotowych modeli rozpoznających zarówno rękopisy, jak i teksty drukowane, jednak żaden z nich nie nadawał się do wykorzystania przy transliterowaniu *Moralioń*. Wybierając model, należy bowiem zwrócić uwagę na kilka parametrów. Oczywiście, najistotniejsze jest bliskie wizualne podobieństwo między tekstem, który chcemy automatycznie transliterować, a tekstem spożytkowanym do wytrenowania modelu. W praktyce oznacza to, że model do kursywy angielskiej (*Copperplate*) nie zadziała na próbce zapisanej uncją – analogicznie: wystarczy, aby charakter pisma naszego autora delikatnie różnił się od charakteru pisma z modelu, a efekty automatycznej transliteracji nie będą już tak zadowalające. Poza tą oczywistą kwestią warto też pamiętać o wskaźniku CER (*Character Error Rate*), który określa, jak często komputer myli się, odczytując znaki; im wyższy CER (powyżej 5%), tym mniej dokładny odczyt. CER 15% oznacza około piętnastu źle rozpoznanych znaków na sto, czyli prawie piętnaście korekt na standardową linijkę tekstu dwunastopunktowego czcionką Times New Roman w programie Microsoft Word – to bardzo dużo, dlatego poprawianie takiej transkrypcji może pochłoniąć więcej czasu niż stworzenie jej od zera.

Zastanawiając się nad wyborem gotowego rozwiązania, należy także uwzględnić język tekstu, na którym AI się szkoliło. Model wytrenowany na anglojęzycznych dokumentach z pewnością nie rozpozna polskich znaków diakrytycznych, będzie też częściej popełniał błędy, ponieważ nie przyswoił sobie typowych dla języka polskiego, a nieobecnych w angielskim zestawień znaków. Więcej – obserwacje transkrypcji generowanych w Transkribusie dowodzą, że model uczy się nie tylko znaków, lecz także kształtu całych wyrazów, dlatego do pewnego stopnia jest w stanie poprawiać tekst skanów (!)³³. W Transkribusie niewiele publicznych modeli nadaje się do wykorzystania przez polskich edytorów. W galerii gotowych rozwiązań zawierającej ponad dwieście modeli – od cerkiewnosłowiańskiego po kursywę tybetańską – tylko trzy są przeznaczone do języka polskiego³⁴. Spośród nich dwa

33 Nie zawsze jest to pożądane. Przykładowo model „The Polish Schwabacher” przeznaczony do generowania transliteracji polskiej szwabachy konsekwentnie kreskuje a pochylone w słowie *náprzod* (*naprzód*), nawet gdy na skanie wyraźnie widać literę a bez kreski. Edytor, który w transliteracji chciałby zachować tę chwiejną pisownię, musi zachować też... czujność.

34 Stan na grudzień 2024 r.

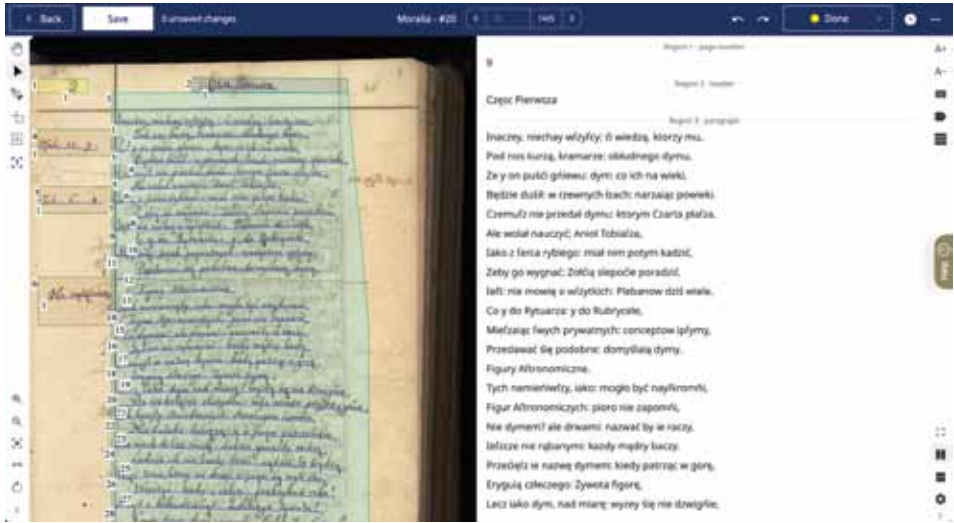
to stale rozwijane duże modele „dokarmiane” dużymi zbiorami danych – jeden do druku, drugi do rękopisów (wielojęzyczny Transkribus Print M1 – CER 2,2%, Transkribus Polish M2 – CER 4,1%). Trzeci model, The Polish Schwabacher (CER 0,87%), został opracowany na początku 2024 roku przez studentki edytorstwa na Uniwersytecie Jagiellońskim w ramach zajęć z tworzenia cyfrowych edycji naukowych i służy do automatycznego transliterowania szwabachy.

Dla osiągnięcia najlepszych możliwych rezultatów należało wytrenować nowy model specjalnie przeznaczony do odczytywania rękopisów Potockiego. Wiązało się to z koniecznością przygotowania próbki rzeczywistych danych (*ground truth*), które algorytm potraktowałby jako idealny wzór do naśladowania – innymi słowy: nawet chcąc zautomatyzować pracę, najpierw trzeba było ją w niemałym zakresie samodzielnie wykonać. Transliterację pierwszych stu kart rękopisu sporządziła Lidia Nowak, absolwentka edytorstwa, obecnie doktorantka w Szkole Doktorskiej Nauk Humanistycznych UJ. Odczytanie manuskryptu musiało być jak najwierniejsze i możliwie pozbawione dwuznaczności, ponieważ w próbie *ground truth* każdy znak pisma z zabytku powinien mieć swój stały odpowiednik w transliteracji – w przeciwnym razie trening sztucznej inteligencji nie będzie aż tak skuteczny. Dla komputera każdy znak jest tak samo abstrakcyjny i pozbawiony znaczenia: jeżeli konsekwentnie będziemy mu pokazywać, że wielka litera A na skanie jest tak naprawdę małym g, zacznie ją w ten sposób rozpoznawać.

Zgodnie z wcześniej przyjętymi zasadami transliteracja *Moraliów* między innymi:

- zachowała układ i łamanie stron, wersów oraz marginaliów,
- zachowała znaki przestankowania występujące w przekazie (bez zachowania odstępów niekonsekwentnie poprzedzających te znaki),
- rozwijała ligatury *æ, œ, &* do *ae, oe i et*,
- ujednoliciła trzy warianty grafemu *z, z, ʒ* do *z* (analogicznie ze znakiem *ž* – jeżeli warianty miały kropkę u góry),
- długie *s* oddawała przez *ʃ*,
- zachowała decyzje autora dotyczące użycia wielkich liter (w tym niekonsekwentne, ale wyraźne rozróżnienie majuskułowych liter *I* oraz *J*),
- rozwijała abrewiatyry,
- za pomocą nawiasów *< >* oznaczała miejsca uszkodzone i nieczytelne.

Transliteracja została wykonana w edytorze tekstu Word, dlatego należało ją przenieść do Transkribusa i powiązać ze skanami zabytku. Procedura przebiegała w kilku etapach. Najpierw pobrano skany *Moraliów* udostępnione przez Cyfrową Bibliotekę Narodową „Polona”, następnie poprzez sekwencję wsadową (funkcja programu Photoshop) skany rozkładówek zostały pomniejszone oraz podzielone



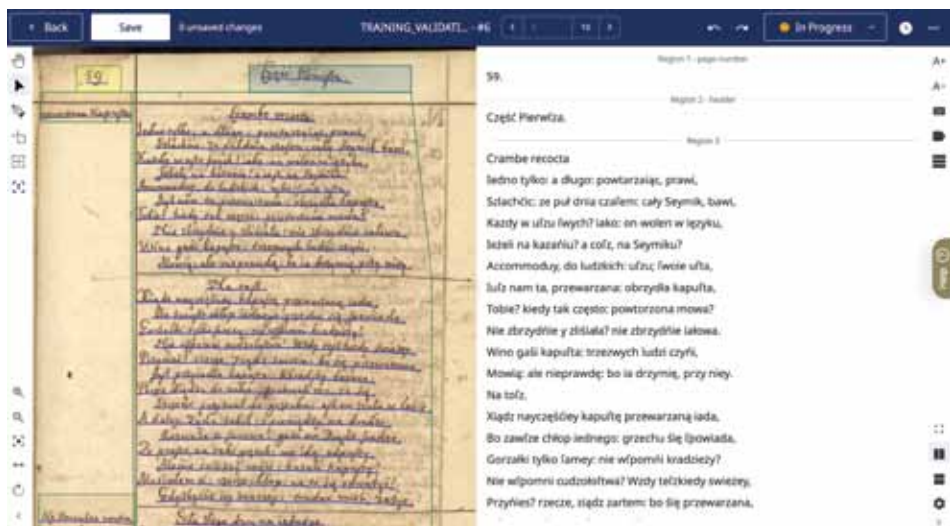
Il. 4. Aplikacja przeglądarkowa Transkribus – podgląd na faksymile i transliterację. Na stronach rękopisu poza tekstem głównym wyróżniono pola tekstowe z paginacją, marginaliami oraz żywą paginą.

nie będą później odczytywane. Na etapie budowania layoutu wiele zależy od jakości skanu. Jeżeli na przykład digitalizowana książka nie rozwierała się w pełni, to wewnętrzna część kolumny na skanie będzie się „zawijała” ku linii grzbietu. Takie zawinięcia sprawiają dwojaką trudność Transkribusowi: najpierw przy rozpoznaniu layoutu, później przy odczytaniu znaków (gdyż de facto są one zniekształcone i pochylone – nie bez przyczyny obrazki CAPTCHA wyświetlane na niektórych stronach internetowych i w elektronicznych formularzach jako zabezpieczenie przed botami zawierają „falujące” litery, które człowiek rozpozna bez problemu, a komputer niekoniecznie... przynajmniej do niedawna).

Gotowa transliteracja *Moralioń* została przyporządkowana do uprzednio zmapowanych pól tekstowych i linijek tekstu. Oczywiście, Transkribus umożliwia wpisywanie transliteracji bezpośrednio w oknie przeglądarki, jednak wcześniejsze sporządzenie próbki w Wordzie ma swoje zalety, związane przede wszystkim z bardziej zaawansowanymi opcjami edycji tekstu dostępnymi w tym edytorze (przeszukiwaniem, zamienianiem, zastosowaniem wyrażen regularnych); nie bez znaczenia jest także stabilność działania, z którą aplikacja webowa Transkribusa wciąż miewa problemy.

Twórcy usługi sugerują w dokumentacji, że do wytrenowania efektywnego modelu wystarczy próbka danych zawierająca od pięciu do piętnastu tysięcy słów³⁶. Modele rozpoznające tekst drukowany zwykle radzą sobie dobrze nawet na małych

36 Transkribus Help Center, 2. *Data Preparation*, <https://help.transkribus.org/data-preparation>.



Il. 5. Zestaw walidacyjny dla modelu Potocki-2. Z prawej strony tekst odczytany przez komputer, brak ingerencji edytora.

próbkach, rękopisy potrzebują dużo większego materiału. Pierwotnie zgodnie z instrukcją zespół LabEdytu zamierzał wykorzystać pięćdziesiąt stron manuskryptu (około piętnastu tysięcy słów), jednak model uzyskany z tej próbki osiągnął dość wysoki CER – 5,9%. Dopiero po wprowadzeniu dodatkowych pięćdziesięciu stron transliteracji udało się obniżyć CER do satysfakcjonującego poziomu 2,6%.

Jakie wnioski płyną z tego eksperymentu? Przede wszystkim należy się zastanowić, kiedy tak naprawdę warto zainwestować czas w szkolenie modelu AI do rozpoznawania tekstu. Gdyby rękopis *Moralików* liczył trzysta stron, być może szybciej, łatwiej i dokładniej byłoby go własnoręcznie przepisać do pliku tekstowego. Wykorzystanie sztucznej inteligencji w dalszym ciągu przecież nie zwalnia wydawcy naukowego z obowiązku uważnej, kilkukrotnej korekty tekstu; samo przygotowanie próbki do treningu również jest czasochłonne. Mimo to przedstawiona tu procedura ma dość dużą przewagę, jeśli w dalszej perspektywie zamierzamy stworzyć cyfrową edycję naukową. Przyjazny dla użytkownika interfejs graficzny Transkribusa został wyposażony w użyteczne funkcje tagowania, dzięki którym już na etapie transkrypcji (czy też transliteracji – w zależności od potrzeb użytkownika) można oznaczać miejsca nieczytelne, uszkodzone, skreślone, uzupełnione lub skorygowane, a ponadto wskazywać elementy struktury przekazu źródłowego, jak na przykład paginację, żywą paginę, kustosze, sygnatury, marginalia, tytuły i tym podobne. Od niedawna istnieje możliwość stosowania tagów <persName> oraz <placeName>, jak również definiowania własnych znaczników.

Wszystko to można uzyskać, nie wprowadzając ani jednej linijki kodu, co stanowi niewątpliwą zaletę dla tych edytorów, którzy niechętnie korzystają z języka znaczników. Pozyskany w ten sposób tekst w rzeczywistości nie jest czystym tekstem (choć i taki da się z Transkribusa pobrać, ponieważ aplikacja umożliwia eksport w wielu różnych formatach), lecz ma dodatkową warstwę danych, które znajdują się później w wyeksportowanym pliku Page XML konwertowanym do TEI XML³⁷. To znacząco ułatwia pracę nad cyfrową edycją naukową, gdyż pozwala w półautomatyczny sposób uzyskać wstępnie otagowane oraz zmapowane pliki, w których wiersze tekstu czy nawet pojedyncze słowa mają przyporządkowane koordynaty na cyfrowym faksymile.

Obecnie dobiega końca korekta transliteracji całych *Moralioń*. Wkrótce rozpocznie się kolejny etap pracy – półautomatyczne tworzenie transkrypcji modernizującej pisownię. W tym celu zespół zamierza wykorzystać listę frekwencyjną słów uzyskaną dzięki przetworzeniu czystego pliku tekstowego za pomocą nieskomplikowanego skryptu napisanego w Pythonie. W czasie gdy trwały prace nad odczytaniem rękopisu Potockiego (model Potocki-2 powstał 16 czerwca 2023 roku), dokonał się znaczący postęp w tej dziedzinie. Sam Transkribus przeszedł ogromną przemianę i z prostego, właściwie bezpłatnego narzędzia oferującego wówczas bezkonkurencyjne funkcjonalności wyewoluował do komercyjnej usługi działającej w modelu *freemium*³⁸. Warto jednak pamiętać, że cyfrowe edytorstwo naukowe ma dokładnie ten sam cel co edytorstwo tradycyjne – jest nim naukowe opracowanie tekstu zgodnie z przyjętymi założeniami. Dlatego nie należy przesadnie fetyszyzować narzędzi humanistyki cyfrowej, ani też szczególnie się do nich przywiązywać. Jeśli dziś lub wkrótce pojawi się program lepiej spełniający swoje zadanie, a już słyhać opinie³⁹, że może to być udostępniane na otwartej licencji eScriptorium, to nic nie będzie stało na przeszkodzie, by kolejne transkrypcje dzieł literackich powstawały właśnie w nim.

*

W latach sześćdziesiątych ubiegłego wieku szwajcarski typograf Adrian Frutiger we współpracy z europejskim stowarzyszeniem producentów komputerów (ECMA)

37 Transkribus umożliwia eksport danych bezpośrednio do pliku TEI, jednak jest to opcja dostępna wyłącznie dla użytkowników subskrybujących plan premium.

38 Dynamiczne zmiany w dostępie do oprogramowania są istotnym zagrożeniem dla projektów finansowanych w modelu grantowym. Zmuszają one bowiem do uwzględnienia w kosztorysie nie tylko różnicy kursu walut, lecz także potencjalnej podwyżki ceny licencji.

39 Twórcy narzędzia do publikowania naukowych edycji cyfrowych TEI Publisher zapowiadają integrację właśnie z eScriptorium. Oprogramowanie to ma jednak istotną wadę – ze względu na to, że jest rozwijane w modelu *open source*, okazuje się nie tak przyjazne w obsłudze jak konkurencyjny Transkribus.

pracował nad krojem pisma, który byłby rozpoznawalny przez czytniki optyczne. Tak powstał „znormalizowany alfabet łaćniński” OCR-B (*Optical Character Recognition – font B*), będący kompromisem między wymaganiami ówczesnych cyfrowych maszyn a wielowiekową tradycją typografii⁴⁰. Projektant starał się tak narysować poszczególne znaki cyfr i alfabetu, aby żaden z nich (na przykład I i l czy B i 8) nie był podobny do siebie, gdyż „ułamne” komputery nie radziły sobie z drobnymi różnicami kształtu. Frutiger już wtedy nie miał wątpliwości, iż „pewnego dnia maszyna czytająca zostanie tak dalece ulepszona, że będzie mogła bezbłędnie odczytać znaki i formy jakiegokolwiek stylu współczesnych alfabetów”⁴¹. Pracę nad krojem OCR-B, niezbyt urodziwym, typograf uznawał za „sukces w dziedzinie etyki”, ponieważ „to nie maszyna zmusza człowieka do stosowania «zmechanizowanego» stylu znaków, tylko człowiek stara się «nauczyć» maszynę odczytywania pisma będącego w powszechnym zastosowaniu, tego pisma, które przez wieki rozwijało się od kamiennych hieroglifów, poprzez pióro i pergamin, rylec stemplacza, aż po obecne metody grafików literników, składaczy i drukarzy naszych czasów”⁴². Czyżby?

Obserwowana dziś rewolucja sztucznej inteligencji to przyszłość ze snów Frutigera. Jej przebieg stawia wiele pytań natury etycznej oraz egzystencjalnej: o przyszłość dotychczasowych profesji, o ekologię (superkomputery zużywają ogromne ilości energii), o prywatność danych. Te pytania zadają sobie także edytorzy naukowi. Automatyczne generowanie transkrypcji czy transliteracji z wykorzystaniem AI wydaje się zabiegiem dość niewinnym. To oczywiste, że na końcu tekst syczyta i poprawi „instancja ludzka”. Co jednak z próbami tworzenia komentarza przez czatboty? Kto będzie autorem takiej edycji, kto weźmie za nią odpowiedzialność – edytor piszący prompty czy komputer przetwarzający wiedzę całej ludzkości? Czy tekst napisany na nasze polecenie jest naszym tekstem? Czy edytor dalej będzie edytorem – tym, który najgłębiej poznał dzieło – a może już... wyspecjalizowanym operatorem maszyny?

BIBLIOGRAFIA

Brückner A., *Wacława Potockiego „Moralia” (1688)*, wyd. Tadeusz Grabowski, Jan Łoś, [Kraków] 1915–1918 [recenzja], „Pamiętnik Literacki” 1920, t. 17/18, nr 1–4.

C21 Editions. Scholarly Editing and Publishing in the Digital Age, <https://www.c21editions.org/> (wszystkie przywoływane źródła internetowe – stan z 20 września 2024 r.).

Corpus Thomisticum, <https://www.corpusthomicum.org/>.

40 A. Frutiger, *OCR-B. Znormalizowane pismo o czytelności optycznej*, tłum. R. Tomaszewski, „Litera” [dod. „Poligrafiki”] 1968, R. 3, nr 23–25, s. 91–93 [dokończ. w nr. 25, s. 97–102].

41 Ibidem, s. 102.

42 Ibidem.

- DataReportal, *Digital 2024 Global Overview Report*, <https://datareportal.com/reports/digital-2024-global-overview-report>.
- Digital Humanities Lab, <https://dhlab.id.uj.edu.pl/>.
- Eisenstein E.L., *Rewolucja Gutenberga*, tłum. H. Hollender, Warszawa 2004.
- Elektroniczny korpus tekstów polskich XVII i XVIII w. „KorBa”, <https://korba.edu.pl>.
- Frutiger A., *OCR-B. Znormalizowane pismo o czytelności optycznej*, tłum. R. Tomaszewski, „Litera” [dod. „Poligrafiki”] 1968, R. 3, nr 23–25 [dokończ. w nr. 25].
- Gruchała J., *Wacław Potocki – problem edytorski*, w: *Wacław Potocki. W 400-lecie urodzin poety*, red. K. Koehler, D. Chemperek, Warszawa 2023.
- Grzeszczuk S., *O potrzebie i programie badań nad twórczością Wacława Potockiego*, w: *Wśród zagadnień polskiej literatury barokowej*, cz. 2: *Motywy – inspiracje – recepcja*, red. Z.J. Nowak, Katowice 1980.
- Grześkowiak R., *Stary druk jako podstawa edycji krytycznej. Preliminaria*, w: *Jak wydawać teksty dawne*, red. K. Borowiec et al., Poznań 2017.
- Hollender H., *Czy świat czeka przyszłość średniowiecza?*, w: E.L. Eisenstein, *Rewolucja Gutenberga*, tłum. H. Hollender, Warszawa 2004.
- Kukulski L., *Prolegomena filologiczne do twórczości Wacława Potockiego*, Wrocław 1962.
- Pierazzo E., *What Future for Digital Scholarly Editions? From Haute Couture to Prêt-à-Porter*, „International Journal of Digital Humanities” 2019, nr 13.
- Potocki W., *Dziela*, t. 3: *Moralia i inne utwory z lat 1688–1696*, oprac. L. Kukulski, Warszawa 1987.
- Moralia*, rękopis, ok. 1688–1696, Biblioteka Narodowa, rps 3049 III, Polona.pl.
- Wacława Potockiego „Moralia” 1688*, t. 1–3, wyd. T. Grabowski, J. Łoś, Kraków 1915–1918.
- Potocki Wacław (1621–1696)*, w: *Bibliografia literatury polskiej. „Nowy Korbut”*, red. K. Budzyk, t. 3: *Pis-miennictwo staropolskie. Hasła osobowe N–Ż*, oprac. R. Pollak et al., Warszawa 1965.
- Program of the ESTS Conference 2024*, <https://elte-dh.hu/ests-2024-program/>.
- Roterodamus E., *Adagiorum Chiliades Des. Erasmi Roterodami Quatvor Cvm Dimidia Ex Postrema Avtoris Recognitione: In hac aeditione, prioribus tribus Indicibus subiunctus est quartus nouus [...]*, <https://rcin.org.pl/dlibra/publication/74578/edition/66983/content>.
- Shillingsburg P.L., *Od Gutenberga do Google’a. Elektroniczne reprezentacje tekstów literackich*, tłum. P. Bem, Warszawa 2020.
- Silberling A., *Why AI Can’t Spell ‘Strawberry’*, <https://techcrunch.com/2024/08/27/why-ai-cant-spell-strawberry/>.
- Tong A., Paul K., *Exclusive: OpenAI Working on New Reasoning Technology under Code Name ‘Strawberry’*, <https://www.reuters.com/technology/artificial-intelligence/openai-working-new-reasoning-technology-under-code-name-strawberry-2024-07-12>.
- Transkribus – Unlocking the Past with AI, <https://www.transkribus.org/>.
- Viglianti R., del Rio Riande G., *Against Infrastructure. Global Approaches to Digital Scholarly Editing*, „C21 Digital Editions” 2023.
- Viglianti R., del Rio Riande G., Hernández N., De León R., *Otwarte, minimalne i promujące równe szanse. Jak uczy się tworzenia naukowych edycji cyfrowych na Globalnej Północy i Globalnym Południu*, tłum. J. Wełniak, „Sztuka Edycji” 2023, nr 1.

SŁOWA KLUCZE: Wacław Potocki (1621–1696), cyfrowa edycja naukowa, zastosowanie sztucznej inteligencji w humanistyce, humanistyka cyfrowa, automatyczne rozpoznawanie tekstu

THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN SCHOLARLY EDITING: A CASE STUDY OF WACŁAW POTOCKI'S *MORALIA*

In the early 2020s, Artificial Intelligence has become extremely popular. Due to new technology, more processes from various fields are subject to automation every day. Editing is no different, as concepts of utilising the feats of engineering to create scholarly editions are voiced ever more confidently. In the international community of Humanists, there is an ongoing debate on the possibilities for creative uses of generative artificial intelligence and machine learning in the work with text.

The first part of the article portrays the current state of scholarly editing in the digital age. The second part is devoted to a case study of utilising machine learning for the automatic generation of a transliteration of the seventeenth-century manuscript of *Moralia* by Wacław Potocki. The text concludes with a discussion on the ethics of editing aided by AI. The endeavour is realised at the Jagiellonian University as part of the Digital Humanities Lab project.

KEYWORDS: Wacław Potocki (1621–1696), digital scholarly edition, the application of AI in the Humanities, Digital Humanities, OCR and HTR