NAUKA 2/2026 • 89–106 DOI: 10.24425/nauka.2026.2297

Łukasz Remisiewicz *

Sztuczna inteligencja a recenzje naukowe: modele wykorzystania, polityki i wyzwania

Wprowadzenie

Postęp w dziedzinie sztucznej inteligencji w ostatnich latach wywiera rosnący wpływ na wiele aspektów komunikacji naukowej, w tym na proces recenzowania publikacji naukowych. Pojawienie się dużych modeli językowych (LLM) pokroju ChatGPT, Gemini, Grok czy Claude otworzyło nowe możliwości usprawnienia pracy recenzentów, ale jednocześnie wzbudziło poważne kontrowersje. Z jednej strony AI obiecuje wsparcie – może automatycznie korygować język recenzji, streszczać obszerne manuskrypty czy nawet sugerować pytania i komentarze merytoryczne. Z drugiej strony pojawiają się obawy o rzetelność i transparentność procesu recenzji, gdy nie wiadomo, czy otrzymana opinia została napisana przez człowieka, czy wygenerowana przez model AI. Niektórzy autorzy przyznają, że demotywuje ich podejrzenie, iż ich praca została oceniona „przez robota" zamiast przez eksperta. Te obawy potwierdzają pierwsze incydenty: w 2023 r. wykryto przypadki recenzji grantów i artykułów, które były częściowo lub w całości wygenerowane przez ChatGPT. Przykładowo na czołowych konferencjach z dziedziny sztucznej inteligencji oszacowano, że do 17% złożonych ostatnio recenzji zawierało istotne fragmenty tekstu wygenerowane przez modele językowe. Jedna z takich recenzji chwaliła artykuł za „klarowną narrację i logicznie uporządkowane sekcje", co autor pracy uznał za absurdalnie brzmiącą, schematyczną pochwałę – łatwo kojarzącą się ze stylem ChatGPT (Grove, 2023; Lee, 2024; Liang et al., 2024b; Zou, 2024).

Równolegle ze względu na rosnącą produktywność naukową pojawia się coraz więcej wyzwań związanych z funkcjonowaniem samego systemu recenzji. Warto wspomnieć tu choćby o zjawisku *reviewer fatigue* – zmęczenia i przeciążenia recenzentów – które staje się poważnym problemem w obliczu eksplozji liczby publikacji naukowych. Rocznie ukazuje się już ponad 3 miliony artykułów, co przerasta możliwości uważnej lektury przez ludzi. Coraz trudniej jest pozyskać kompetentnych recenzentów, a ci, którzy podejmują się zadania, są przeciążeni i często działają pod presją czasu. W części analiz odsetek zaproszeń, które nie kończą się uzyskaniem recenzji, przekracza 60%. W tej sytuacji narzędzia AI jawią się jako potencjalna deska ratunku – mogłyby automatyzować część rutynowych zadań i odciążyć ludzi. Już od kilku lat wydawnictwa wykorzystują

* Dr Łukasz Remisiewicz (lukasz.remisiewicz@ug.edu.pl), Uniwersytet Gdański, Instytut Socjologii

elementy AI do celów takich jak wyszukiwanie recenzentów pasujących do tematyki manuskryptu czy wstępna kontrola antyplagiatowa. Teraz, wraz z pojawieniem się zaawansowanych modeli generatywnych, pojawia się pokusa, by AI angażować bezpośrednio w pisanie recenzji lub formułowanie rekomendacji publikacyjnych (Fox et al., 2017; National Science Board, 2023; Hosseini i Horbach, 2023).

Powyższe zjawiska rodzą zasadnicze pytania: czy i jak integrować AI w proces recenzowania naukowego? Społeczność akademicka jest w tej kwestii podzielona. Niektórzy postulują daleko posuniętą ostrożność lub wręcz zakazy użycia AI przy recenzjach – argumentując, że recenzja powinna pozostać domeną ludzkiej oceny i odpowiedzialności. Inni wskazują, że odpowiednio użyta AI może poprawić efektywność i obiektywizm recenzji, o ile zachowane są zasady przejrzystości. W efekcie wydawcy i organizacje naukowe zaczynają wydawać oficjalne wytyczne definiujące dopuszczalne granice użycia AI przez recenzentów. Celem niniejszego artykułu jest przegląd tych dynamicznych zmian – od praktycznych zastosowań AI jako narzędzia pomocniczego po normy etyczne i regulacje – oraz ocena, czy AI stanowi zagrożenie dla integralności procesu recenzji, czy też szansę na jego usprawnienie (Hosseini i Horbach, 2023; Li et al., 2024).

W kolejnych częściach omawiam: (1) wykorzystanie AI do poprawy języka i stylu recenzji, (2) zastosowanie AI we wspieraniu merytorycznej oceny prac, (3) wyzwania etyczne i ryzyka związane z AI w recenzowaniu (w tym odpowiedzialność, uprzedzenia, przejrzystość i nadużycia), (4) aktualne stanowiska i polityki wydawców oraz instytucji względem AI w recenzjach oraz (5) problem nierówności w dostępie do narzędzi AI i ich wpływ na globalną równowagę w procesie recenzji.

AI jako pomoc językowa dla recenzenta

Jednym z najbardziej oczywistych sposobów wykorzystania sztucznej inteligencji w procesie recenzji jest wsparcie recenzenta w poprawie języka i stylu przygotowywanej opinii. Tradycyjnie recenzenci – zwłaszcza ci recenzujący prace w języku obcym – muszą poświęcać czas na dopracowanie klarowności wypowiedzi, korektę błędów gramatycznych czy stylistycznych. Nowoczesne modele językowe mogą w tym kontekście pełnić rolę zaawansowanych „asystentów redakcyjnych". Przykładowo, ChatGPT czy podobne narzędzia potrafią przepisać zadaną wypowiedź poprawniej i zwięźlej, zachowując sens pierwotnej treści. Recenzent może więc teoretycznie wpisać szkic swojej opinii do takiego systemu, prosząc o poprawienie języka lub uproszczenie zdań. AI może też służyć jako inteligentny korektor – wskazujący literówki, błędy interpunkcyjne czy niejednoznaczne sformułowania (Hosseini i Horbach, 2023).

Zastosowanie AI do korekty językowej ma szczególne znaczenie w międzynarodowym obiegu naukowym, gdzie językiem dominującym jest angielski. Recenzenci z krajów, w których angielski nie jest językiem ojczystym, często zmagają się z barierami

językowymi. Narzędzia takie jak Grammarly czy DeepL już od pewnego czasu wspierają piszących po angielsku w poprawianiu jakości tekstu. Modele generatywne dają jeszcze większe możliwości – potrafią np. przeformułować całe akapity w bardziej naturalny sposób, dostosować ton wypowiedzi (np. uczynić go uprzejmym, lecz stanowczym) czy streścić zbyt rozwlekły fragment. Dla recenzenta oznacza to potencjalnie oszczędność czasu i pewność, że jego uwagi zostaną zrozumiane przez autora tak, jak zamierzał, a nie zagubią się w nie do końca poprawnym języku (Lillis i Curry, 2010; Hosseini i Horbach, 2023).

W praktyce naukowej pojawiły się już przypadki korzystania z AI do celów językowych przy recenzowaniu – choć nie zawsze fortunne. Ben Maier, niemiecki badacz, opisał sytuację, gdy redaktor jednego z czasopism zasugerował użycie ChatGPT do „ulepszenia jasności" tekstu w rękopisie, co *de facto* oznaczało wygenerowanie fragmentów recenzji lub zaleceń edytorskich przez AI. W efekcie powstałe paragrafy brzmiały sztucznie i niepoprawnie stylistycznie, co skłoniło autora do wycofania pracy z procesu publikacji – uznał on bowiem, że wygenerowane przez AI poprawki przekraczały dopuszczalne normy interwencji w tekście. Ten incydent pokazuje, że pochopne zastosowanie AI do celów redakcyjnych może obniżyć jakość recenzji zamiast ją podnieść, zwłaszcza jeśli model wygeneruje sformułowania nieadekwatne do stylu dyskursu naukowego (Grove, 2023).

Kluczowe jest tu rozróżnienie między drobną pomocą językową a całościowym generowaniem tekstu recenzji. Wydawcy naukowi generalnie akceptują tę pierwszą formę wsparcia, o ile nie narusza ona poufności (o czym dalej) i jest ograniczona do stylistycznych poprawek. Przykładowo, Springer Nature stwierdza w swoich wytycznych, że niewielka pomoc AI w redakcji językowej recenzji może być dopuszczalna, jednak recenzent musi ujawnić fakt skorzystania z takiego narzędzia w treści raportu dla redakcji. Podobnie wiele innych wydawnictw przyjmuje, że użycie AI do poprawy „czytelności, gramatyki lub formatowania" tekstu nie wymaga formalnej deklaracji – dotyczy to zwłaszcza autorów prac, ale analogię stosuje się do recenzentów. Z kolei Elsevier przyjmuje bardziej rygorystyczne stanowisko: wprost przestrzega recenzentów, by nie udostępniali treści recenzji żadnym zewnętrznym narzędziom AI (nawet w celu korekty języka), gdyż recenzja zawiera informacje poufne o manuskrypcie. W polityce Elseviera czytamy, że raport recenzencki również podlega zasadzie poufności, więc wprowadzanie go do publicznie dostępnego modelu językowego (np. ChatGPT, Gemini, Grok etc.) jest zabronione, nawet jeśli chodzi tylko o ulepszenie stylu. Wydawca ten obawia się nie tylko wycieku danych, ale i zatarcia odpowiedzialności – przypomina, iż recenzja jest pracą twórczą recenzenta, za którą tylko on może odpowiadać, a „krytyczne myślenie i oryginalna ocena" wymagana w recenzji przekracza możliwości obecnych narzędzi AI (Elsevier, 2025; Nature Portfolio, b.d.; Li et al., 2024).

Warto zauważyć, że istnieje cienka granica między poprawą języka a współautorstwem tekstu recenzji. Jeśli recenzent używa AI jedynie do sugerowania poprawek w zdaniach, pozostając autorem merytorycznej treści, to w istocie AI pełni rolę narzędzia edytorskiego – analogicznie jak sprawdzanie pisowni. Jeśli jednak recenzent polega na AI przy formułowaniu całych akapitów z opinią naukową, to zaciera się granica między tym, co napisał człowiek, a co maszyna. Taka sytuacja rodzi już istotne kwestie etyczne i musi być przedmiotem jawności (jeśli w ogóle miałaby być dopuszczona). Większość wytycznych (np. COPE – Committee on Publication Ethics) kładzie nacisk na to, by recenzenci zawsze ujawniali, czy korzystali z chatbotów w formułowaniu jakichkolwiek fragmentów recenzji (COPE, 2021; COPE, 2023; ICMJE, b.d.).

AI jako wsparcie merytoryczne w ocenie prac

Znacznie bardziej dyskusyjne – aczkolwiek już eksperymentalnie sprawdzane – jest wykorzystanie sztucznej inteligencji do wsparcia merytorycznej strony procesu recenzowania. Modele AI mogą tutaj pełnić szereg ról: od szybkiej analizy treści manuskryptu (streszczenie, wyodrębnienie głównych tez i wyników), poprzez ocenę spójności logicznej wywodu, aż po porównanie pracy z istniejącą literaturą czy wykrywanie potencjalnych słabości metodologicznych. Niektórzy entuzjaści sugerują, że AI mogłaby służyć jako drugi pararecenzent, np. generując wstępną listę uwag, którą ludzki recenzent następnie weryfikuje i rozwija (Hosseini i Horbach, 2023; Liang et al., 2024a).

Już teraz dostępne są narzędzia wykorzystujące AI do oceny jakości tekstu naukowego. Przykładowo, serwisy preprintów eksperymentują z automatyczną oceną przejrzystości raportowania badań czy kompletności bibliografii. Jednak najdalej idące próby to użycie modeli generatywnych (jak GPT-4) do wygenerowania całego raportu recenzji na podstawie dostarczonego manuskryptu. Taki eksperyment przeprowadzili m.in. Marrella i in. (2025), którzy sprawdzili, jak ChatGPT poradzi sobie z recenzowaniem 11 opublikowanych artykułów z chirurgii ręki. Model otrzymał pełne teksty prac (były to już publikacje, co rozwiązało problem poufności) i wygenerował do każdej z nich proponowaną decyzję (przyjąć/odrzucić) oraz uzasadnienie w formie recenzji. Wyniki porównano z oryginalnymi recenzjami ludzkimi – oceniając m.in. jakość i użyteczność komentarzy według standaryzowanej skali (ARCADIA). Okazało się, że pod pewnymi względami AI wypadła zaskakująco dobrze: recenzje wygenerowane przez ChatGPT były dłuższe i bardziej szczegółowe, a ich „jakość redakcyjna" oceniona w skali ARCADIA była wyższa (średnio ok. 4,8/5 punktów) niż jakość recenzji pisanych przez ludzi (2.8–3.2/5). Innymi słowy, model AI potrafił wygenerować obszerne, dobrze zorganizowane raporty, które na pierwszy rzut oka przewyższały skrótowe uwagi wielu recenzentów (Marrella et al., 2025).

Interpretacja tego wyniku wymaga jednak ostrożności. Po pierwsze, ChatGPT cechował się wyraźną tendencją do nadmiernej pobłażliwości: w eksperymencie niemal 95–98% ocen AI sugerowało akceptację artykułu, podczas gdy ludzcy recenzenci w rzeczywistości znaczną część tych prac odrzucali lub wymagali istotnych poprawek. AI brakowało więc krytycyzmu – model zbyt mocno podkreślał mniej ważne aspekty, takie jak poprawność językowa czy ogólna struktura artykułu. Po drugie, szczegółowa analiza wykazała, że zgodność decyzji ChatGPT z decyzjami czasopism o najwyższym prestiżu była bardzo niska (tylko ~30% przypadków), natomiast częściej pokrywała się z linią bardziej tolerancyjnych czasopism o niższym wskaźniku Impact Factor. Sugeruje to, że model rozmijał się z fachowym osądem co do wartości naukowej pracy. Po trzecie, nawet jeśli formalne oceny aspektów pracy (np. poprawność statystyczna) w recenzjach AI były trafne, pozostaje ryzyko „halucynacji", czyli generowania przekonująco brzmiących, lecz fałszywych informacji. Autorzy eksperymentu zauważyli, że konieczne było wydanie modelowi bardzo precyzyjnych instrukcji, by unikał konfabulacji i skupiał się na treści dostarczonego artykułu. Innymi słowy, AI trzeba było „pilnować", aby nie zbaczała poza dane wejściowe – co i tak nie gwarantuje uniknięcia wszystkich błędów (Marrella et al., 2025).

Mimo to, potencjał AI jako pomocnika w ocenie merytorycznej jest realny. LLM potrafi błyskawicznie streścić długi manuskrypt, wyłapując główne punkty – co dla ludzkiego recenzenta może być przydatną weryfikacją, czy niczego nie przeoczył. AI może też służyć do wykrywania pewnych wzorców lub sygnałów alarmowych: np. nierzetelnie wysokiej zgodności z innymi tekstami (plagiat), nienaturalnych danych (np. wygenerowanych losowo), niezgodności między częścią opisową a danymi liczbowymi itp. Co więcej, modele trenowane na olbrzymich korpusach literatury mogą przypominać recenzentowi o ważnych pracach, które autorzy pominięli w bibliografii – niejako podpowiadać kontekst literaturowy. Już obecnie powstają dedykowane systemy, które jako „asystenci recenzenta" integrują różne algorytmy: np. sprawdzające, czy wszystkie cytowane prace istnieją i są prawidłowo przytoczone, analizujące, czy wnioski autora rzeczywiście wynikają z przytoczonych wyników, czy nie ma sygnałów manipulacji obrazami, itp. (Hosseini i Horbach, 2023; Liang et al., 2024a).

Z drugiej strony, nadmierne poleganie na AI w ocenie merytorycznej budzi zrozumiałe opory. Istnieje obawa, że recenzent kuszony łatwością wygenerowania gotowych zdań przestanie samodzielnie myśleć krytycznie o pracy – stanie się tylko „operatorem" przekazującym sugestie modelu. To rodzi pytania w szczególności o utratę kompetencji recenzowania u młodych naukowców: jeśli od początku będą wyręczani przez AI w formułowaniu uwag, być może nigdy nie rozwiną w pełni umiejętności wnikliwej analizy naukowej. Ponadto modele są częściowo nieprzejrzyste, co sprawia, że mogą mieć wbudowane algorytmy faworyzujące pewne dziedziny, metody lub typy wyników, które

dominują w danych treningowych. Taka stronniczość mogłaby przełożyć się na recenzje. Bez pełnej przejrzystości co do mechanizmów modelu trudno zaufać, że jego merytoryczne oceny są neutralne. Zwraca się też uwagę, że AI nie ma „instynktu" naukowca – nie wyczuje intuicyjnie nowatorskości albo potencjalnego wpływu pracy na rozwój dziedziny, co jest ważnym, choć subiektywnym, elementem oceny merytorycznej (Hosseini i Horbach, 2023).

W związku z powyższym AI może być wartościowym wsparciem merytorycznym recenzenta, o ile jest używana jako narzędzie pomocnicze, a nie decydujący arbiter. Praktyka, którą proponują niektórzy, to tzw. recenzja wspomagana AI: najpierw recenzent czyta pracę i wyrabia sobie opinię, następnie może zapytać model (na własnym, bezpiecznym komputerze, bez udostępniania treści pracy publicznej AI) o streszczenie pracy, listę potencjalnych zalet i wad, propozycje pytań do autorów itp. Recenzent traktuje to jako check-listę – analizuje sugestie AI, odrzuca nietrafne (świadcząc tym samym o krytycznym podejściu), ewentualnie korzysta z pewnych spostrzeżeń, o których sam by nie pomyślał. Ostatecznie całość decyzji i sformułowań należy do recenzenta. Takie podejście mogłoby łączyć najlepsze cechy obu stron – maszynowej szybkości i ludzkiego osądu. Jednak wymaga dużej samodyscypliny i świadomości ograniczeń AI. Jak trafnie zauważyli Hosseini i in. (2023), LLM mogą znacznie zmienić rolę recenzenta, ale ich fundamentalna nieprzejrzystość i możliwe błędy rodzą poważne znaki zapytania co do pełnego zaufania takim systemom. Autorzy ci sugerują na razie podejście ostrożne: jeśli korzystać z AI przy pisaniu recenzji, to tylko z pełnym ujawnieniem i wzięciem odpowiedzialności za ewentualne błędy, a także z gotowością na to, że model może wzmacniać istniejące uprzedzenia lub problemy systemu recenzji zamiast je rozwiązywać (Hosseini i Horbach, 2023).

Etyka recenzowania naukowego z użyciem AI

Włączenie AI w proces recenzji stawia szereg dylematów etycznych. Kluczowe z nich dotyczą: odpowiedzialności, przejrzystości, bezstronności (braku uprzedzeń) oraz ryzyka nadużyć. Kwestie te są na tyle istotne, że doczekały się już pierwszych zaleceń ze strony organizacji zajmujących się etyką publikacyjną (COPE, 2021; COPE, 2023; ICMJE, b.d.).

(a) **Odpowiedzialność za treść recenzji.** Fundamentalną zasadą etyczną jest to, że recenzent jest odpowiedzialny za swoją opinię i musi być gotów bronić jej merytorycznej rzetelności. Gdy w grę wchodzi AI, pojawia się pytanie: kto odpowiada za ewentualne błędne lub wprowadzające w błąd uwagi wygenerowane przez model? Jasne stanowisko zajmuje tu m.in. International Committee of Medical Journal Editors (ICMJE), które w aktualnych rekomendacjach podkreśla, iż AI nie może być traktowana jako autor czegokolwiek, co składa się na proces publikacyjny – dotyczy to zarówno pisania

manuskryptu, jak i recenzowania. AI nie spełnia kryteriów autorstwa (nie bierze odpowiedzialności za prawdziwość treści), więc recenzent nie może przypisywać jej „współautorstwa" swoich ocen. W praktyce oznacza to, że jeśli recenzent używa AI, i tak ponosi pełną odpowiedzialność za całą treść recenzji – tak, jakby napisał ją samodzielnie. Z tego powodu wiele wytycznych (Elsevier, Springer, COPE i in.) podkreśla: to recenzent jest autorem raportu recenzenckiego i nawet najdrobniejsze wykorzystanie AI nie zwalnia go z bycia świadomym i kontrolującym każdą część przekazywanej opinii. Innymi słowy, recenzent może użyć narzędzia, ale musi zweryfikować jego odpowiedzi w taki sposób, jakby sprawdzał pracę asystenta – poprawiając błędy, usuwając bezzasadne uwagi, upewniając się co do prawdziwości każdego stwierdzenia. W przeciwnym razie narusza standardy rzetelności (ICMJE, b.d.; Elsevier, 2025; Nature Portfolio, b.d.; COPE, 2023).

(b) **Transparentność**. Etyka akademicka wymaga ujawniania istotnych informacji, które mogą wpływać na ocenę procesu. W kontekście AI oznacza to, że recenzent powinien jawnie poinformować redakcję (a pośrednio i autorów) o każdym niebłahym użyciu AI w przygotowaniu recenzji. Takie zalecenie formułuje choćby COPE, stwierdzając, że recenzenci (podobnie jak autorzy i redaktorzy) powinni deklarować użycie chatbotów lub innych narzędzi AI w trakcie oceny manuskryptu. Podobny kierunek widać w części czasopism medycznych analizowanych przez Li i współautorów oraz w polityce Taylor & Francis, które podkreślają potrzebę jawności i ograniczenia użycia AI do przypadków dopuszczonych przez pismo. Springer Nature wymaga od recenzentów, którzy wspomagali się AI „w ocenie", by zaznaczyli to w swoim raporcie. Taka informacja może np. przybrać formę zdania: „Recenzent informuje, że przy formułowaniu poniższych uwag wspomagał się narzędziem AI (ChatGPT) w zakresie poprawy języka/streszczenia argumentacji, przy zachowaniu poufności danych". Chodzi tutaj, po pierwsze, o to, by utrzymać zaufanie do procesu – autor i redakcja mają prawo wiedzieć, jak powstała recenzja i czy ewentualne sformułowania nie są np. efektem mechanicznej manipulacji. Po drugie, by umożliwić weryfikację – jeśli w recenzji pojawiłyby się jakieś osobliwe błędy (np. halucynacje), redakcja świadoma użycia AI może łatwiej dociec przyczyny i poprosić recenzenta o wyjaśnienia lub poprawki. Brak transparentności mógłby prowadzić do sytuacji, w której recenzje wygenerowane przez AI krążą w obiegu, udając w pełni ludzkie – co byłoby formą oszustwa akademickiego i naruszeniem zasad rzetelności (podobnie jak *ghostwriting*). W skrajnym przypadku zatajenie użycia AI przez recenzenta można traktować jako naruszenie etyki publikacyjnej, analogiczne do nieujawnienia istotnego konfliktu interesów (COPE, 2021; Nature Portfolio, b.d.; Li et al., 2024; Taylor & Francis, b.d.).

(c) **Poufność i prywatność danych**. Proces *peer review* jest z definicji poufny – recenzent otrzymuje nieopublikowany manuskrypt w zaufaniu, że nie będzie rozpow-

szechniał zawartych w nim danych ani wyników przed ich oficjalnym upublicznieniem. Wykorzystanie AI rodzi tu poważne zagrożenie: większość popularnych modeli (jak ChatGPT) to usługi online działające na zewnętrznych serwerach, które zapisują wprowadzone do nich treści i potencjalnie wykorzystują je do dalszego trenowania modelu lub innych celów. Wprowadzenie fragmentu niepublikowanego artykułu do takiego systemu mogłoby oznaczać niekontrolowane ujawnienie treści osobom trzecim. Dlatego najważniejsze polityki wydawnicze i grantowe zakazują wprowadzania treści manuskryptu do otwartych narzędzi AI. Springer Nature wprost zaznacza: recenzenci nie mogą uploadować manuskryptów do ChatGPT czy podobnych platform ze względu na ochronę poufnych informacji i praw autorskich autorów. Podobnie Elsevier przypomina, że rękopis jest dokumentem poufnym i nie wolno go przetwarzać żadnym narzędziem, które mogłoby go zachować lub udostępnić. Nawet ICMJE sugeruje, że jeśli recenzent chciałby użyć AI, powinien uzyskać uprzednią zgodę redakcji – tak by upewnić się, że nie dojdzie do naruszenia zasad poufności czy praw do danych. Jedyną ewentualną furtką jest korzystanie z modeli AI wdrożonych lokalnie lub przez wydawcę, które gwarantują nieprzechwytywanie danych. Niektórzy wydawcy (np. Elsevier) rozwijają własne algorytmy do pomocy recenzentom (np. do sprawdzania spójności piśmiennictwa czy rekomendowania recenzentów), które działają w obrębie zamkniętego systemu i respektują poufność. Jednak używanie publicznie dostępnych chatbotów wprost na treści artykułu przed publikacją jest – z powodów etycznych i prawnych – wykluczone. Warto tu przypomnieć głośny przypadek z Australasian Research Council: wykryto, że w procesie recenzji grantów recenzenci użyli ChatGPT, co natychmiast uznano za naruszenie zasad i doprowadziło do wprowadzenia formalnego zakazu użycia AI w recenzjach w tej instytucji. Podobnie postąpiły agencje finansujące badania w Wielkiej Brytanii, które we wspólnym oświadczeniu podkreśliły, że korzystanie z generatywnej AI przy ocenie wniosków może zagrażać poufności i integralności procesu, w związku z czym nie będzie tolerowane (Elsevier, 2025; Nature Portfolio, b.d.; ICMJE, b.d.; NIH, 2023; Australian Research Council, 2023; UK Research and Innovation, 2025).

(d) **Uprzedzenia i obiektywizm**. Wspominałem już wcześniej, że modele AI mogą kierować się rozmaitymi domyślnymi regułami generującymi stronnicze odpowiedzi. W kontekście recenzji naukowej szczególnie istotny jest problem uprzedzeń epistemicznych. Istnieje ryzyko, że AI może np. faworyzować prace z krajów anglojęzycznych (bo anglojęzyczna literatura dominowała w treningu) lub powielać utarte schematy oceniania wartości naukowej (np. wyżej ceniąc prace z instytucji o znanych nazwach). Jeżeli recenzent bezkrytycznie wykorzystywałby sugestie AI, mógłby nieświadomie przenosić te uprzedzenia do swoich ocen. Co więcej, jak pokazują najnowsze analizy, zastosowanie AI niekoniecznie usuwa istniejące biasy recenzentów – czasem je tylko maskuje lub przenosi w inne obszary. Stanfordzkie badanie (Lepp & Smith, 2025) wykazało, że choć

po pojawieniu się ChatGPT styl językowy prac z całego świata się wyrównuje (gramatyczne niedoskonałości znikają dzięki AI), to recenzenci zaczęli doszukiwać się innych sygnałów świadczących o tym, skąd pochodzi autor, np. rozpoznawali charakterystyczne frazy często generowane przez LLM i kojarzyli je z autorami z krajów nieanglojęzycznych. Innymi słowy, uprzedzenia geograficzno-językowe w recenzjach nie znikły, a tylko zmieniły formę: recenzenci zamiast „słabego angielskiego" zaczęli piętnować „brzmienie jak z AI" – co dalej stygmatyzowało autorów spoza głównego nurtu, bo to oni częściej korzystali z pomocy AI językowo. Okazuje się więc, że istnieje głębszy problem: sama technologia nie zlikwiduje ludzkich uprzedzeń, a może nawet stworzyć nowe (Liang et al., 2023; Lepp i Smith, 2025).

Kolejny przykład uprzedzeń to kwestia płci i pochodzenia recenzowanych autorów. Modele AI trenowane na dotychczasowych danych mogą powielać trend, w którym np. badaczki lub naukowcy z regionów słabiej reprezentowanych otrzymują surowsze oceny. Co prawda pewne badania sugerują, że AI można zaprogramować, by wręcz korygowała te nierówności, np. Teixeira (2025) pokazał, że GPT-4 poproszony o wytypowanie recenzentów do danego artykułu, spontanicznie wybierał w większości mężczyzn z krajów wysokorozwiniętych, ale gdy dodano wytyczne uwzględniające różnorodność, model potrafił przedstawić zrównoważoną pod względem płci i geografii listę ekspertów, nie gorszych merytorycznie. To pozytywny sygnał, że AI może też pomagać przełamywać pewne biasy – o ile używa jej się w sposób świadomy. W przypadku pisania recenzji, można by analogicznie starać się „uwrażliwić" model na różnorodność podejść i kontekstów, by unikał np. eurocentrycznej czy androcentrycznej perspektywy. Jednak taka złożona kontrola w generowaniu tekstu jest jeszcze słabo opanowana i w praktyce wymaga od użytkownika świadomej, aktywnej interwencji na każdym etapie pracy z modelem. W praktyce więc społeczność naukowa nakłada na recenzenta obowiązek bycia strażnikiem obiektywizmu – jeśli korzysta z AI, musi krytycznie ocenić jej sugestie pod kątem potencjalnych uprzedzeń i je skorygować (Teixeira, 2025).

(e) **Potencjalne nadużycia i oszustwa**. Niestety, wykorzystanie AI w recenzowaniu stwarza pole do nowych form nieetycznych zachowań. Już odnotowano incydenty, gdzie recenzent „posiłkujący się" AI generował szkodliwe lub nieuczciwe treści. Głośny stał się przypadek recenzenta, który odrzucając artykuł, zasugerował autorowi zapoznanie się z szeregiem prac – rzekomo ważnych przeglądów literatury – których autorzy i tytuły brzmiały wiarygodnie, ale okazały się fikcyjne. Analiza wykazała, że prawdopodobnie ChatGPT został poproszony o wygenerowanie listy „dodatkowych lektur" i „zhalucynował" on nieistniejące źródła. Recenzent najwyraźniej bezrefleksyjnie wkleił taką listę do swojej opinii, w efekcie wprowadzając autora w błąd i łamiąc podstawowy obowiązek recenzenta, jakim jest rzetelność. Emerald Publishing, wydawca tego czasopisma, uznał sytuację za na tyle poważną, że publicznie potępił użycie AI w ten sposób

i przypomniał, iż recenzja ma opierać się na ekspertyzie człowieka, a nie generatywnych fantazjach modelu. Tego typu działanie nazwano wręcz sabotowaniem procesu recenzji. Trudno się nie zgodzić – generowanie fałszywych informacji jest sprzeczne z etyką, a w kontekście recenzji może wyrządzić realną szkodę (autor może tracić czas na szukanie nieistniejących źródeł, podważa się też zaufanie do recenzji) (Grove, 2023).

Szczególnym typem nadużycia jest próba napisania recenzji bez przeczytania pracy. Taka sytuacja jest oczywiście trudna do udowodnienia, ale pewne symptomy mogą ją zdradzić: recenzje odrealnione, zbyt ogólnikowe i nieodnoszące się konkretnie do treści artykułu. Redakcje już teraz stają przed wyzwaniem wykrywania wygenerowanych recenzji. Pojawiają się narzędzia, takie jak detektory AI, którymi można próbować ocenić, czy dany tekst ma charakter syntetyczny, jednak nie są one w 100% niezawodne. Dlatego silny nacisk kładzie się na świadomość redaktorów – polityki Nature Portfolio oraz komentarze dotyczące praktyki recenzenckiej podkreślają potrzebę uważnej lektury napływających recenzji pod kątem nietypowego języka, halucynacji i braku związku z treścią manuskryptu. Jeden z redaktorów przyznał, że jeśli dostaje recenzję, która jest bardzo krótka i „oderwana" od tematu, poważnie rozważa poproszenie innego recenzenta o ocenę. Na razie nie ma znanych przypadków oficjalnych sankcji za nieuprawnione, a w szczególności nieetyczne użycie AI, dyskusje na temat granic odpowiedzialności w tym względzie wciąż trwają (Nature Portfolio, b.d.; Liang et al., 2023; Lee, 2024).

Wytyczne i polityki wydawnicze dotyczące AI w *peer review*

W reakcji na gwałtowny rozwój narzędzi generatywnych, w latach 2023–2025 wiele czołowych wydawnictw oraz instytucji regulujących etykę publikacji opracowało oficjalne polityki określające dopuszczalne użycie AI przez autorów, recenzentów i redaktorów. Poniżej przeanalizuję stanowiska wybranych organizacji. Warto zaznaczyć, że wytyczne te są dość spójne co do zasad nadrzędnych, różnią się jednak szczegółami i stopniem restrykcyjności (Li et al., 2024).

Elsevier. Ten wydawca przyjął jedno z najbardziej restrykcyjnych podejść wobec użycia AI w recenzjach. W opublikowanej we wrześniu 2025 r. polityce jednoznacznie stwierdzono, że generatywna AI nie powinna być używana przez recenzentów do naukowej oceny pracy. Elsevier argumentuje, że krytyczna analiza i ocena manuskryptu to zadania wymagające ludzkiej ekspertyzy i nie mogą być delegowane maszynie, która może wygenerować błędne lub stronnicze wnioski. W szczególności zabrania się wgrywania tekstu recenzowanego artykułu do jakichkolwiek narzędzi AI (kwestia poufności, o której była mowa). Co więcej, wydawca podkreśla, że nawet raport recenzencki jest dokumentem poufnym i także nie powinien być umieszczany w AI do celów językowej korekty. Elsevier dopuszcza właściwie tylko korzystanie z własnych, wewnętrznych

narzędzi AI (spełniających standardy poufności RELX) do wsparcia recenzentów i redaktorów, np. systemów do wyszukiwania odpowiednich recenzentów czy sprawdzania plagiatów (Elsevier, 2025).

Springer Nature. Polityka tego wydawcy jest bardziej otwarta, choć również wyraźnie podkreśla rolę człowieka jako decydującą. Według opublikowanych wskazówek recenzenci nie mogą udostępniać manuskryptu generatywnej AI ze względu na poufność i ochronę praw autorów. Jeśli natomiast jakakolwiek część oceny roszczeń przedstawionych w manuskrypcie była wsparta przez narzędzie AI, recenzent powinien ujawnić to w raporcie recenzenckim. Springer Nature wskazuje więc, że AI może co najwyżej wspierać pracę recenzenta, ale nie zastępuje osądu eksperckiego ani odpowiedzialności za treść opinii. Samo Nature Portfolio w polityce redakcyjnej podkreśla również, że zasady te będą aktualizowane wraz z rozwojem technologii i praktyk społeczności naukowej (Nature Portfolio, b.d.).

COPE (Committee on Publication Ethics). COPE jako wiodąca organizacja etyki publikacyjnej wydała wprawdzie bardziej ogólne zalecenia, ale jednoznacznie wskazała pewne standardy. W swoich dokumentach dotyczących AI COPE podkreśla, że uczestnicy procesu wydawniczego powinni jasno określać, czy i w jaki sposób użyli narzędzi AI, a także że odpowiedzialność za ocenę i treść raportu pozostaje po stronie człowieka. Zasada ta zbiega się z omówioną wcześniej transparentnością. Ponadto COPE zajęło stanowisko, że AI nie może być autorem – co przekłada się na recenzje o tyle, że nie wolno wymieniać AI jako recenzenta, ale też recenzent nie powinien traktować modelu jako podmiotu odpowiedzialnego za ocenę. COPE zwraca także uwagę na potrzebę szkolenia i zwiększania świadomości recenzentów co do możliwości i ograniczeń AI, tak aby mogli podejmować świadome decyzje o jej ewentualnym użyciu. Wydawcy często odwołują się do wytycznych COPE, kształtując własne regulacje dotyczące jawności, odpowiedzialności i uczciwości w erze AI (COPE, 2021; COPE, 2023).

ICMJE. Międzynarodowy Komitet Redaktorów Czasopism Medycznych zaktualizował w 2023 r. swoje rekomendacje, by uwzględnić AI. W odniesieniu do recenzentów ICMJE zaznacza, że użycie AI przez recenzenta może naruszać poufność, dlatego każdorazowo recenzent powinien skonsultować się z redakcją, zanim skorzysta z takiego narzędzia. W praktyce to dość silne zastrzeżenie: recenzent musi zapytać o pozwolenie, co stawia spory próg formalny – można sądzić, że większość redakcji i tak by odradziła stosowania takich praktyk. ICMJE podkreśla też, że recenzenci – podobnie jak autorzy – powinni uważać na fakt, że tekst generowany przez AI bywa błędny, niepełny lub stronniczy (ICMJE, b.d.).

Warto zauważyć, że poza modelami skrajnie restrykcyjnymi i warunkowo dopuszczającymi istnieje też grupa polityk pośrednich. Taylor & Francis nie pozwala recenzentom używać generatywnej AI do tworzenia raportów recenzenckich, ale dopuszcza jej

wykorzystanie do poprawy języka gotowej recenzji; recenzent pozostaje przy tym w pełni odpowiedzialny za dokładność i integralność swojego raportu. Z kolei wydawca czasopism Science (AAAS) ogłosił, że AI nie może być używana do generowania ani pisania recenzji. W obu przypadkach wspólnym mianownikiem pozostają poufność manuskryptu oraz prymat ludzkiego osądu w *peer review* (Taylor & Francis, b.d.; AAAS, 2023).

Poza wydawcami, również organizacje finansujące badania odnoszą się do tej kwestii. NIH w USA, ARC w Australii i UK Research and Innovation w Wielkiej Brytanii wprowadziły zakazy lub bardzo rygorystyczne ograniczenia użycia generatywnej AI w recenzowaniu wniosków grantowych. Motywowane jest to przede wszystkim ochroną poufności, integralności procesu oraz ryzykiem, że modele wygenerują błędne, niepełne lub stronnicze oceny. Można oczekiwać, że w miarę dojrzewania dyskusji pewne zalecenia będą się dalej ujednolicać, ale już dziś widać wyraźny wspólny rdzeń tych polityk: AI nie może przejmować eksperckiej odpowiedzialności za ocenę (NIH, 2023; Australian Research Council, 2023; UK Research and Innovation, 2025).

Nierówności epistemiczne w dostępie do narzędzi AI

Na koniec warto omówić często pomijany aspekt, jakim są nierówności w dostępie do AI oraz wynikające z nich konsekwencje dla globalnego systemu recenzowania. Narzędzia AI – zwłaszcza najnowocześniejsze modele językowe – nie są równomiernie dostępne dla wszystkich badaczy i recenzentów na świecie. Różnice mają charakter zarówno językowy, jak i infrastrukturalny, co przekłada się na potencjalne nierówności epistemiczne: nie każdy recenzent ma takie same możliwości skorzystania z AI, by usprawnić swoją pracę, a to może pogłębiać istniejące dysproporcje w systemie naukowym (Besiroglu et al., 2024; Hammerschmidt et al., 2025).

Infrastruktura i koszty. Modele generatywne wymagają sporych zasobów – mocnych komputerów lub dostępu do chmury oraz szybkiego internetu. W wielu rejonach świata recenzenci nie dysponują taką infrastrukturą. Nawet jeśli model jest otwarty, potrzeba odpowiedniego sprzętu, by go uruchomić lokalnie. Alternatywą jest dostęp przez internet, ale czołowe modele są komercyjne i płatne. W bogatych instytucjach można nie tylko wykupić dostęp do najwyższych modeli, ale także tworzyć własne. Natomiast w słabiej dofinansowanych uczelniach zakup dostępu do AI może nie być priorytetem, a pojedynczych naukowców często po prostu nie stać na samodzielne finansowanie takiego narzędzia. Powstaje ryzyko, że recenzowanie wspomagane AI stanie się luksusem dostępnym głównie w krajach najbogatszych, co może zwiększyć dysproporcje w efektywności pracy naukowców. Ci z dostępem mogą szybciej i sprawniej wykonywać recenzje (co może przełożyć się np. na to, że będą bardziej pożądanymi recenzentami, dostaną więcej zleceń, zbudują lepsze CV naukowe), podczas gdy inni będą odstawać (Besiroglu et al., 2024; Hammerschmidt et al., 2025).

Biasy kulturowe modeli a recenzenci spoza głównego nurtu. Nawet zakładając, że wszyscy mieliby równy dostęp do infrastruktury technicznej, dochodzi kwestia, czy AI jednakowo wspiera różnych recenzentów. Modele tworzone i trenowane są głównie w kontekstach zachodnich, więc mogą nie uwzględniać pewnych lokalnych uwarunkowań ważnych np. w humanistyce czy naukach społecznych w innych kulturach. Recenzent z określonego obszaru badań (np. etnolog badający lokalną społeczność) może uznać sugestie AI za nieprzystające do kontekstu. Jeśli jednak redakcja – dysponując AI – zacznie bardziej ufać ujednoliconym, „sformalizowanym" recenzjom (które AI ułatwia produkować), to głosy odbiegające stylem lub punktem widzenia (często recenzentów z innych tradycji akademickich) mogą być mniej cenione. Już się obserwuje, że AI może homogenizować styl recenzji – badania recenzji konferencyjnych wskazują na efekt ujednolicania tekstu wraz ze wzrostem udziału treści modyfikowanych przez LLM. Jeśli taki „ujednolicony" styl zostanie uznany za normę, to recenzenci o odmiennym podejściu mogą być postrzegani jako odstający. To może zrodzić subtelne formy nierówności epistemicznych w ramach promowania jednego wzorca oceny wiedzy (tego, który AI łatwo reprodukuje) kosztem pluralizmu stylów i perspektyw (Liang et al., 2024b; Couldry i Mejias, 2019).

Język a postrzeganie jakości naukowej. Wspomniane badanie Lepp & Smith (2025) wykazało, że recenzenci mogą wyrażać swoje uprzedzenia względem naukowców z krajów peryferyjnych, np. doszukując się „nienaturalnego brzmienia". To pokazuje, że w sytuacji, gdy „wszyscy piszą poprawnie", naturalnie wytwarzane są nowe formy tych samych podziałów. W skrajnym przypadku używanie AI może samo stać się piętnowane jako „oznaka słabości językowej", co dotknie nieproporcjonalnie naukowców z grup niedominujących językowo. Badacze wskazują, że potrzeba tu zmiany podejścia recenzentów, a nie tylko uzbrojenia autorów/recenzentów w AI (Lepp i Smith, 2025; Lillis i Curry, 2010).

Demokratyzacja czy pogłębienie przepaści? Zwolennicy AI często głoszą, że będzie ona „demokratyzującą" technologią – wyrównującą szanse, bo np. każdy z dostępem do internetu może mieć „asystenta". Rzeczywistość już teraz okazuje się bardziej złożona. Jak zauważają krytycy, sama dostępność technologii nie gwarantuje sprawiedliwości – potrzebne są też zmiany instytucjonalne i uwrażliwienie społeczności. Bez tego AI może pogłębić nierówności: ci, którzy już są uprzywilejowani (lepsza uczelnia, lepsza infrastruktura), wyciągną z niej najwięcej korzyści, zwiększając swoją przewagę. Jednocześnie grupy marginalizowane mogą znaleźć się w sytuacji, gdzie oczekuje się od nich efektywności wspomaganej AI, ale nie zapewnia im się realnie dostępu czy szkoleń. Istnieje też ryzyko, że kraje rozwijające się staną się poligonem dla różnych automatycznych systemów recenzyjnych bez właściwego nadzoru, co mogłoby obniżyć jakość tamtejszych procesów publikacyjnych, podczas gdy elitarne czasopisma dalej będą stawiać

na drobiazgową pracę ekspertów (Hammerschmidt et al., 2025; Zhang et al., 2024; Cobo et al., 2024).

Co można z tym zrobić? Przede wszystkim, dyskutując o implementacji AI w recenzowaniu, należy uwzględniać perspektywę globalną. UNESCO w swojej Rekomendacji dot. Otwartej Nauki (2021) podkreślała konieczność zmniejszania barier dla naukowców z krajów rozwijających się – dotyczy to także dostępu do narzędzi cyfrowych. Być może dobrym kierunkiem byłoby opracowanie otwartych modeli AI dedykowanych wspieraniu recenzji, które byłyby dostępne za darmo dla środowiska akademickiego na całym świecie (podobnie jak obecnie otwarte repozytoria czy oprogramowanie typu Open Journal Systems). To jednak wymaga międzynarodowej współpracy i funduszy. Ponadto, świadomość uprzedzeń musi być elementem szkoleń recenzentów: recenzenci powinni być uczuleni, że AI może nieświadomie popychać ich ku decyzjom reprodukującym *status quo* (UNESCO, 2021; Hammerschmidt et al., 2025; Cobo et al., 2024).

Nierówności epistemiczne wydają się jednym z najpoważniejszych wyzwań w erze AI. Jeśli nic się nie zmieni, istnieje ryzyko, że AI stanie się kolejnym czynnikiem dzielącym naukę na „centrum" i „peryferie", gdzie centrum dysponuje najnowszymi narzędziami i narzuca standardy, a peryferie starają się nadążyć. Aby temu przeciwdziałać, potrzebna jest świadoma polityka: zapewnienie szerszego dostępu do narzędzi (np. licencje edukacyjne, tańsze wersje dla biedniejszych regionów), rozwój lokalnych modeli i – przede wszystkim – kultura recenzowania oparta na inkluzyjności. Recenzent wspierany AI powinien być tak samo wyczulony na kontekst i różnorodność, jak recenzent tradycyjny – jeśli nie bardziej (Besiroglu et al., 2024; Couldry i Mejias, 2019; UNESCO, 2021).

Zakończenie

Rozwój sztucznej inteligencji stawia społeczność naukową przed koniecznością przemyślenia utrwalonych praktyk recenzowania. Z jednej strony narzędzia AI oferują atrakcyjne możliwości: mogą przyspieszyć pracę recenzenta, wyłapać błędy, usprawnić komunikację i częściowo złagodzić problem przeciążenia systemu *peer review*. Z drugiej strony niekontrolowane użycie AI grozi podważeniem zaufania do recenzji poprzez wprowadzanie niejawnych automatyzmów, błędów czy uprzedzeń (Fox et al., 2017; Hosseini i Horbach, 2023).

W niniejszym artykule starałem się przedstawić najważniejsze obszary tej złożonej problematyki. AI jako pomoc językowa okazuje się względnie mało kontrowersyjna: przy zachowaniu poufności i odpowiedzialności modele mogą skutecznie poprawiać styl i klarowność recenzji, co jest cenne zwłaszcza w międzynarodowym środowisku wielojęzycznym. AI jako wsparcie merytoryczne budzi więcej emocji – eksperymenty pokazują, że potrafi generować rozbudowane recenzje, ale brakuje jej ludzkiej przenikliwości i może

wykazywać systematyczne błędy oceny. Dlatego zalecany jest tu model pracy hybrydowej: sprzężenie zalet AI z krytycznym nadzorem człowieka, zamiast zastępowania recenzenta. Aspekty etyczne – odpowiedzialność, transparentność, unikanie uprzedzeń – zostały już częściowo uregulowane w oficjalnych wytycznych. Konsensus wydaje się jasny: recenzent nie może abdykować na rzecz AI, a poufność i uczciwość nie podlegają negocjacjom; tam, gdzie AI wspiera samą ocenę, polityki albo wymagają ujawnienia jej użycia, albo po prostu tego zabraniają. Analiza polityk wydawniczych pokazuje, że większość renomowanych graczy (Elsevier, Springer Nature, Nature, ICMJE, COPE i inni) opublikowało zbieżne w duchu reguły: AI nie może być „recenzentem-widmo", ale dopuszcza się ograniczoną pomoc z zachowaniem ostrożności i jawności (Marrella et al., 2025; Liang et al., 2024a; Elsevier, 2025; Nature Portfolio, b.d.; ICMJE, b.d.).

Nie można jednak zapominać o szerszym kontekście. Wykorzystanie AI w recenzowaniu dzieje się w realiach globalnych nierówności w dostępie do wiedzy. Bezrefleksyjne wdrożenie AI grozi ryzykiem, że bogate instytucje i dominujące społeczności naukowe jeszcze umocnią swoją pozycję, podczas gdy inni zostaną w tyle – tworząc nową przepaść cyfrową w nauce. Dlatego warto zadbać o to, by demokratyzacja dostępu do narzędzi AI stała się częścią dyskusji. Otwarta nauka powinna oznaczać nie tylko otwarte dane i publikacje, ale i otwarte narzędzia – tak by recenzent z dowolnego kraju miał szansę skorzystać z dobrodziejstw AI na równych zasadach. W przeciwnym razie AI może niechcący utrwalić właśnie te nierówności, które świat nauki stara się redukować (Besiroglu et al., 2024; Couldry i Mejias, 2019; UNESCO, 2021).

Na koniec muszę dodać, że tekst pisany jest w marcu 2026 r. Dynamiczny rozwój modeli sztucznej inteligencji niewątpliwie sprawi, że po pewnym czasie wnioski tu przedstawiane przynajmniej częściowo się zdezaktualizują. Zachodzą bowiem równolegle dwa procesy – z jednej strony modele LLM są cały czas udoskonalane, przekraczając kolejne granice i możliwości. To, co jeszcze dwa lata temu było dla nich nieosiągalne, teraz bez trudu mogą wykonać. Z drugiej strony zachodzą procesy społeczne, które można określić przede wszystkim jako praktyki adaptacji do tej nowej technologii (i jej kolejnych wcieleń). Można zakładać, że i tu w niedługim czasie pojawią się nowe konfiguracje.

Bibliografia

AAAS (2023) Change to policy on the use of generative AI and large language models. Available at: https://www.science.org/content/blog-post/change-policy-use-generative-ai-and-large-language-models (Dostęp: 3 marca 2026).

Australian Research Council (2023) Policy on Use of Generative Artificial Intelligence in the ARC's grants programs. Available at: https://www.arc.gov.au/sites/default/files/2023-07/Policy%20on%20Use%20of%20Generative%20Artificial%20Intelligence%20in%20the%20ARCs%20grants%20programs%202023.pdf (Dostęp: 3 marca 2026).

Besiroglu T., Bergerson S.A., Michael A., Heim L., Luo X., Thompson N. (2024) 'The Compute Divide in Machine Learning: A Threat to Academic Contribution and Scrutiny?', arXiv:2401.02452.

Cobo C., Muñoz-Najar A., Bertrand M. (2024) 100 Student Voices on AI and Education. Washington, DC: World Bank.

COPE (2021) Artificial intelligence (AI) in decision making. Available at: https://publicationethics.org/guidance/discussion-document/artificial-intelligence-ai-decision-making (Dostęp: 3 marca 2026).

COPE (2023) Authorship and AI tools. Available at: https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools (Dostęp: 3 marca 2026).

Couldry N., Mejias U.A. (2019) The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism. Stanford, CA: Stanford University Press.

Elsevier (2025) Generative AI policies for journals. Available at: https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals (Dostęp: 3 marca 2026).

Fox C.W., Albert A.Y.K., Vines T.H. (2017) 'Recruitment of reviewers is becoming harder at some journals: a test of the influence of reviewer fatigue at six journals in ecology and evolution', Research Integrity and Peer Review, 2, 3. doi: 10.1186/s41073-017-0027-x.

Grove J. (2023) 'ChatGPT-generated reading list' sparks AI peer review debate. Times Higher Education, 5 April.

Hammerschmidt T., Stolz K., Posegga O. (2025) 'Bridging the gap: inequalities that divide those who can and cannot create sustainable outcomes with AI', Behaviour & Information Technology. doi: 10.1080/0144929X.2025.2500451.

Hosseini M., Horbach S.P.J.M. (2023) 'Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review', Research Integrity and Peer Review, 8(1), 4. doi: 10.1186/s41073-023-00133-5.

ICMJE (b.d.) AI use by reviewers. Available at: https://www.icmje.org/recommendations/browse/artificial-intelligence/ai-use-by-reviewers.html (Dostęp: 3 marca 2026).

Lee S.M. (2024) AI Scientists Have a Problem: AI Bots Are Reviewing Their Work. The Chronicle of Higher Education, 21 August.

Lepp H., Smith D.S. (2025) '"You Cannot Sound Like GPT": Signs of language discrimination and resistance in computer science publishing', in Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. doi: 10.1145/3715275.3732202.

Li Z.-Q., Xu H.-L., Cao H.-J., Liu Z.-L., Fei Y.-T., Liu J.-P. (2024) 'Use of Artificial Intelligence in Peer Review Among Top 100 Medical Journals', JAMA Network Open, 7(12), e2448609. doi: 10.1001/jamanetworkopen.2024.48609.

Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J. (2023) 'GPT detectors are biased against non-native English writers', Patterns, 4(7), 100779. doi: 10.1016/j.patter.2023.100779.

Liang W. et al. (2024a) 'Can Large Language Models Provide Useful Feedback on Research Papers? A Large-Scale Empirical Analysis', NEJM AI, 1(8). doi: 10.1056/AIoa2400196.

Liang W. et al. (2024b) 'Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews', in Proceedings of the 41st International Conference on Machine Learning. PMLR 235, s. 29575–29620.

Lillis T.M., Curry M.J. (2010) Academic Writing in a Global Context: The Politics and Practices of Publishing in English. London: Routledge.

Marrella D., Jiang S., Ipaktchi K., Liverneaux P. (2025) 'Comparing AI-generated and human

peer reviews: A study on 11 articles', Hand Surgery and Rehabilitation, 44(4), 102225. doi: 10.1016/j.hansur.2025.102225.
National Science Board (2023) Publications Output: U.S. Trends and International Comparisons 2022. Available at: https://ncses.nsf.gov/pubs/nsb202333 (Dostęp: 3 marca 2026).
Nature Portfolio (b.d.) Artificial Intelligence (AI) – editorial policies. Adres internetowy: https://ww.nature.com/nature-portfolio/editorial-policies/ai (Dostęp: 3 marca 2026).
NIH (2023) NOT-OD-23-149: The Use of Generative Artificial Intelligence Technologies is Prohibited for the NIH Peer Review Process. Adres internetowy: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-23-149.html (Dostęp: 3 marca 2026).
Taylor & Francis (b.d.) AI Policy. Adres internetowy: https://taylorandfrancis.com/our-policies/ai-policy/ (Dostęp: 11 marca 2026).
Teixeira A.L. (2025) 'AI in peer review: can artificial intelligence be an ally in reducing gender and geographical gaps in peer review? A randomized trial', Research Integrity and Peer Review, 10(1), 23. doi: 10.1186/s41073-025-00182-y.
UK Research and Innovation (2025) Assessment of applications. Adres internetowy: https://ww.kri.org/councils/mrc/guidance-for-reviewers/peer-reviews/carrying-out-a-peer-review/assessment-criteria/ (Dostęp: 3 marca 2026).
UNESCO (2021) UNESCO Recommendation on Open Science. Adres internetowy: https://n.unesco.org/science-sustainable-future/open-science/recommendation (Dostęp: 3 marca 2026).
Zhang C. (Xinyi), Rice R.E., Wang L.H. (2024) 'College students' literacy, ChatGPT activities, educational outcomes, and trust from a digital divide perspective', New Media & Society. doi: 10.1177/14614448241301741.
Zou J. (2024) 'ChatGPT is transforming peer review – how can we use it responsibly?', Nature, 635(8037), p. 10. doi: 10.1038/d41586-024-03588-8.

Sztuczna inteligencja a recenzje naukowe: modele wykorzystania, polityki i wyzwania

Celem artykułu jest krytyczna analiza najnowszych trendów wykorzystania sztucznej inteligencji (AI) w procesie recenzowania naukowego. Omówiono zarówno korzyści, jak i zagrożenia związane z zastosowaniem narzędzi AI przez recenzentów. W pierwszej kolejności przedstawiono AI jako pomoc językową – np. do korekty stylistycznej i przeformułowania treści recenzji. Następnie przeanalizowano AI jako wsparcie merytoryczne w ocenie prac, w tym możliwości i ograniczenia systemów AI w formułowaniu wniosków i wykrywaniu problemów naukowych. W dalszej części omówiono kwestie etyczne, takie jak odpowiedzialność recenzenta za treści powstałe przy udziale AI, konieczność przejrzystości co do użycia tych narzędzi, ryzyko uprzedzeń zakodowanych w modelach oraz potencjalne nadużycia (np. generowanie fałszywych cytowań). W kolejnym segmencie zreferowano aktualne wytyczne czołowych wydawców i instytucji (m.in. Elsevier, Springer Nature, COPE, Nature, ICMJE) dotyczące korzystania z AI przez recenzentów, ze szczególnym uwzględnieniem zasad poufności i obowiązku ujawniania użycia AI. Na koniec omówiono problem nierówności epistemicznych – w szczególności barier językowych i infrastrukturalnych – w dostępie recenzentów z różnych regionów do zaawansowanych narzędzi AI.

Słowa kluczowe: AI, technologie cyfrowe, recenzje naukowe, *peer review*, nierówności w nauce

Artificial intelligence and scientific peer review: usage models, policies, and challenges

The aim of this article is to critically analyze the latest trends in the use of artificial intelligence (AI) within the scientific peer review process. Both the benefits and risks associated with the application of AI tools by reviewers are discussed. First, AI is presented as a linguistic aid—for example, in stylistic proofreading and rewording review content. Next, AI is analyzed as substantive support in evaluating manuscripts, including the capabilities and limitations of AI systems in formulating conclusions and detecting scientific problems. Subsequently, ethical issues are explored, such as the reviewer's responsibility for AI-assisted content, the need for transparency regarding the use of these tools, the risk of biases encoded in the models, and potential abuses (e.g. generating fake citations). The following segment outlines the current guidelines of leading publishers and institutions (including Elsevier, Springer Nature, COPE, Nature, and ICMJE) regarding the use of AI by reviewers, with a particular focus on confidentiality principles and the obligation to disclose AI usage. Finally, the problem of epistemic inequalities is addressed – specifically linguistic and infrastructural barriers – affecting the access of reviewers from different regions to advanced AI tools.

Key words: AI, digital technologies, scientific reviews, peer review, inequalities in science