

dr hab. Jacek Stańdo, prof. PŁ

Politechnika Łódzka

Polskie Towarzystwo Diagnostyki Edukacyjnej

dr Anna Kreft

Pomorski Ośrodek Doskonalenia Nauczycieli w Słupsku

Poza zwycięstwem: nowe podejście do oceniania rozwiązań konkursowych z wykorzystaniem sztucznej inteligencji

Wprowadzenie

Poza zwycięstwem oznacza zmianę sposobu postrzegania konkursów matematycznych – z wydarzeń służących wyłącznie wyłanianiu najlepszych uczestników na działania wspierające rozwój ucznia. W takim ujęciu wartością konkursu staje się nie tylko końcowy wynik, lecz także proces rozwiązywania zadań, popełniane błędy, refleksja nad własnym myśleniem oraz informacja zwrotna pomagająca doskonalić strategie rozwiązywania problemów. Współczesne narzędzia sztucznej inteligencji oraz podejście oparte na różnych modelach oceniania umożliwiają realizację tej idei (Hashemi, 2024). Dzięki generowaniu spersonalizowanej informacji zwrotnej, wskazywaniu mocnych stron i obszarów wymagających dalszej pracy oraz proponowaniu kolejnych zadań dostosowanych do poziomu ucznia konkurs może stać się elementem procesu uczenia się, a nie jedynie jednorazową formą rywalizacji. Tak rozumiane konkursy matematyczne rozwijają nie tylko kompetencje matematyczne, ale też umiejętność refleksji nad własnym uczeniem się oraz motywację do dalszego rozwoju – niezależnie od zajętego miejsca w rankingu.

Słowa kluczowe: ocenianie rozwiązań matematycznych, konkursy matematyczne, kompetencje matematyczne, adaptacyjne schematy oceniania

Schematy oceniania w konkursach matematycznych z wykorzystaniem sztucznej inteligencji

Ocenianie stanowi integralny element procesu dydaktycznego i pełni znacznie szerszą funkcję niż jedynie klasyfikowanie osiągnięć uczniów. W edukacji matematycznej pozwala nie tylko określić poziom opanowania wiedzy i umiejętności, lecz także dostarcza informacji o sposobie rozumowania ucznia, stosowanych strategiach rozwiązywania problemów oraz najczęściej popełnianych błędach. W literaturze wyróżnia się przede wszystkim ocenianie sumujące, którego celem jest podsumowanie osiągnięć ucznia, oraz ocenianie kształtujące, ukierunkowane na wspieranie procesu uczenia się poprzez udzielanie informacji zwrotnej (Szyling, 2011; Pintał, 2022). Coraz większe znaczenie zyskują również modele oceniania analitycznego, holistycznego i ocenianie efektów uczenia się, umożliwiające bardziej kompleksową ocenę procesu rozwiązywania zadań. Ocena rozwiązań zadań konkursowych jest szczególnym

wyzwaniem, ponieważ oprócz poprawności końcowego wyniku powinna uwzględniać poprawność rozumowania, dobór metod oraz umiejętność uzasadniania kolejnych etapów rozwiązania.

Rozwój narzędzi sztucznej inteligencji stwarza możliwość automatycznego generowania różnorodnych schematów oceniania dostosowanych do wyznaczonego celu oraz oczekiwanego rodzaju informacji zwrotnej. Poniżej zamieszczono pracę ucznia z konkursu matematycznego Pomorskiej Ligi Edukacyjnej, organizowanego w roku szkolnym 2025/2026, która posłuży jako materiał ilustrujący procesy myślowe towarzyszące rozwiązaniu zadań.

Zadanie. Alicja była wtedy o 25 lat młodsza, gdy byłem 2 razy starszy od niej. Razem mamy 98 lat. Ile lat ma Alicja? (*Pomorska Liga Edukacyjna – konkurs matematyczny organizowany w roku szkolnym 2025/2026*).

Przykładowe rozwiązanie ucznia przedstawione jest na rysunku 1.

$$\left\{ \begin{array}{l} \text{Mój wiek} + 25 = 50 \\ \text{wiek Alicji} = 50 - 25 = 25 \\ \text{razem} = 50 + 25 = 75 \end{array} \right.$$

$$\begin{array}{r} 98 \\ - 75 \\ \hline 23 \end{array} \quad \text{czyli } 23 : 2 = 11,5$$

wiec

$$\text{wiek Alicji} + 11,5 = 36,5$$

$$\bullet 50 + 11,5 = 61,5$$

Sprawdzenie:

$$\begin{array}{r} \times \quad 11,5 \\ 50 \quad 25 \\ + 11,5 \quad + 11,5 \\ \hline 61,5 \quad 36,5 \\ \hline 98 \end{array} \quad 61,5 + 36,5 = 98 \checkmark$$

Odp. Alicja ma 36,5 lat.

Rysunek 1. Przykładowe błędne rozwiązanie ucznia

Ocena rozwiązań zadań tekstowych od wielu lat stanowi istotny przedmiot dyskusji w dydaktyce matematyki (Stańdo, 2019). Problem ten wiąże się z pytaniem o cele oceniania oraz o to, jakie kompetencje matematyczne powinny być przez nauczyciela doceniane i rozwijane. W praktyce szkolnej nauczyciele często stają przed dylematem, czy oceniać wyłącznie poprawność otrzymanego wyniku końcowego, czy też brać pod uwagę cały proces dochodzenia do rozwiązania. Zadania tekstowe wymagają bowiem nie tylko wykonania odpowiednich obliczeń, ale również zrozumienia treści, wyodrębnienia istotnych informacji, poprawnego zamodelowania sytuacji matematycznej oraz wyboru właściwej strategii postępowania.

Zwolennicy tradycyjnego podejścia do oceniania podkreślają, że wynik jest najważniejszym kryterium oceny, ponieważ świadczy o skuteczności rozwiązania problemu. Z drugiej strony coraz częściej wskazuje się, że analizowanie samego wyniku nie pozwala w pełni ocenić poziomu wiedzy i umiejętności ucznia. Uczeń może bowiem popełnić drobny błąd rachunkowy, mimo że

poprawnie zrozumiał problem i zastosował właściwą metodę rozwiązania. Podobnie zdarza się, że prawidłowy wynik zostaje uzyskany przypadkowo lub przy użyciu niepełnego rozumowania, co niekoniecznie świadczy o rzeczywistym opanowaniu wymaganych kompetencji.

Przedstawiona praca ucznia stanowi przykład sytuacji, w której rozwiązanie prowadzi do błędnego wyniku, jednak zawiera zauważalne elementy aktywności matematycznej. Uczeń podejmuje próbę analizy zadania, wyodrębnia dane oraz wykonuje działania zmierzające do uzyskania odpowiedzi. Widoczny jest zatem proces myślowy i podejmowanie określonych decyzji matematycznych, nawet jeśli nie wszystkie z nich okazują się trafne. Tego rodzaju aktywność może być interpretowana jako przejaw częściowego zrozumienia problemu i rozwijających się umiejętności rozwiązywania zadań.

W konsekwencji ocena takiej pracy może być różna w zależności od przyjętego modelu oceniania. W modelu nastawionym przede wszystkim na efekt ostateczny uczeń otrzymałby ocenę niską lub brak punktów z uwagi na niepoprawną odpowiedź. Natomiast w podejściu uwzględniającym proces rozwiązywania możliwe byłoby przyznanie części punktów za poprawnie wykonane etapy pracy, trafnie dobraną metodę lub właściwą interpretację niektórych danych. Takie rozwiązanie pozwala lepiej odzwierciedlić rzeczywisty poziom kompetencji ucznia oraz dostarcza mu informacji zwrotnej dotyczącej mocnych i słabych stron jego działania.

Dyskusja nad sposobami oceniania zadań tekstowych odzwierciedla szersze zmiany zachodzące we współczesnej edukacji (Jaksender, Zakrzewski, 2016), w której coraz większe znaczenie przypisuje się nie tylko poprawnym obliczeniom, lecz także rozumowaniu, argumentowaniu i umiejętności samodzielnego rozwiązywania problemów. Dlatego analiza pracy ucznia powinna obejmować zarówno rezultat ostateczny, jak i drogę prowadzącą do jego uzyskania, co pozwala na bardziej sprawiedliwą i diagnostyczną ocenę osiągnięć uczącego się.

W celu przeanalizowania różnych sposobów oceniania rozwiązań zadań tekstowych wykorzystano narzędzie AI (Stańdo i in., 2025), które umożliwiło wygenerowanie kilku alternatywnych schematów oceniania. Dzięki temu możliwe było porównanie podejść koncentrujących się wyłącznie na wyniku końcowym z modelami uwzględniającymi także tok rozumowania ucznia, poprawność zastosowanej metody oraz stopień zrozumienia treści zadania. Wygenerowane schematy pokazały, że ta sama praca ucznia może zostać oceniona w różny sposób w zależności od przyjętych kryteriów. Pozwoliło to lepiej dostrzec zalety i ograniczenia poszczególnych schematów oceniania oraz wskazać, które z nich najpełniej odzwierciedlają rzeczywiste kompetencje matematyczne ucznia.

Ocenianie wyłącznie wyniku końcowego

Najbardziej tradycyjnym sposobem oceniania jest koncentracja na odpowiedzi końcowej. W takim modelu rozwiązanie uznaje się za poprawne wyłącznie wtedy, gdy uczeń poda prawidłowy wynik i spełni wszystkie warunki zadania. Ocena ma charakter zero-jedynkowy: odpowiedź poprawna oznacza otrzymanie pełnej liczby punktów (np. 1 punkt), natomiast odpowiedź błędna skutkuje brakiem punktów (0 punktów).

W analizowanej pracy uczeń podał odpowiedź, że Alicja ma 36,5 roku. Ponieważ poprawnym wynikiem jest 41 lat, rozwiązanie otrzymałoby 0 punktów. Taki sposób oceniania cechuje się wysokim stopniem obiektywności i jest łatwy do zastosowania w przypadku dużej liczby prac. Jednocześnie nie pozwala docenić częściowo poprawnych działań ani wysiłku ucznia. W rezultacie uczeń, który wykonał większość pracy poprawnie, może uzyskać taki sam wynik jak osoba, która nie podjęła żadnej próby rozwiązania.

Należy zwrócić uwagę na to, że w przypadku oceniania odpowiedzi w systemie zero-jedynkowym, gdzie odpowiedź prawidłowa otrzymuje 1 punkt, a błędna 0 punktów, sam wynik końcowy nie zawsze odzwierciedla rzeczywisty sposób rozumowania osoby rozwiązującej zadanie. Może bowiem zdarzyć się sytuacja, w której uczeń popełni błędy na poszczególnych etapach rozwiązania, ale mimo to przypadkowo uzyska prawidłowy wynik końcowy. Przykładowo dwa popełnione błędy mogą się wzajemnie zniwelować, prowadząc ostatecznie do poprawnej odpowiedzi. W takim przypadku nie można uznać zadania za rozwiązane prawidłowo i przyznajemy 0 punktów.

Ocenianie etapów rozwiązania

Znacznie częściej w praktyce szkolnej stosuje się ocenianie polegające na przyznawaniu punktów za kolejne etapy rozwiązania. Zakłada się wówczas, że zadanie można podzielić na kilka kluczowych czynności prowadzących do uzyskania odpowiedzi.

W przypadku omawianego zadania można wyróżnić na przykład następujące etapy:

1. Poprawne zrozumienie informacji zawartej w treści zadania.
2. Utworzenie zależności matematycznej opisującej lata bohaterów.
3. Wykorzystanie informacji o sumie wieku wynoszącej 98 lat.
4. Otrzymanie poprawnego wyniku końcowego.

Analiza pracy ucznia pokazuje, że błędnie zinterpretowano kluczową informację dotyczącą czasu sprzed 25 lat. W konsekwencji nie została zapisana właściwa zależność między latami. Uczeń poprawnie wykorzystał natomiast warunek mówiący o sumie wieku oraz przeprowadził odpowiednie rachunki. W takim modelu oceniania można przyznać częściowe punkty za poprawnie wykonane elementy rozwiązania, nawet jeśli końcowy wynik jest nieprawidłowy.

Zaletą tego podejścia jest możliwość bardziej sprawiedliwej oceny rzeczywistych umiejętności ucznia. Nauczyciel może wskazać dokładnie, na którym etapie pojawił się błąd, a uczeń otrzymuje informację pozwalającą na poprawę własnego sposobu myślenia.

Ocenianie kompetencji matematycznych

Współczesna dydaktyka matematyki coraz częściej odchodzi od oceniania pojedynczego wyniku na rzecz oceny szerszych kompetencji matematycznych. W takim modelu nauczyciel analizuje różne aspekty aktywności ucznia.

Pierwszym z nich jest umiejętność analizy tekstu matematycznego i wydobywania z niego istotnych informacji. Drugim jest tworzenie modelu

matematycznego, czyli przekładanie sytuacji opisanej słownie na działania matematyczne i równania. Trzeci element obejmuje sprawność rachunkową, natomiast czwarty umiejętność kontrolowania poprawności otrzymanego wyniku.

W analizowanej pracy można zauważyć wyraźny kontrast między poszczególnymi kompetencjami. Uczeń poprawnie wykonuje rachunki i podejmuje próbę sprawdzenia odpowiedzi. Jednocześnie nie radzi sobie z właściwą interpretacją tekstu i błędnie modeluje sytuację matematyczną. Ocena kompetencyjna pozwala więc dostrzec, że problem nie wynika z braków rachunkowych, lecz z trudności w rozumieniu zależności opisanych w zadaniu.

Taki sposób oceniania jest szczególnie wartościowy diagnostycznie, ponieważ umożliwia nauczycielowi dokładniejsze określenie obszaru wymagającego dalszej pracy.

Model egzaminacyjny (holistyczny)

W egzaminach zewnętrznych bardzo często stosuje się uproszczone schematy punktowania. Rozwiązanie nie jest wówczas oceniane krok po kroku, lecz przypisywane do określonego poziomu wykonania zadania. Punkty przyznaje się za stopień zbliżenia do poprawnego rozwiązania.

Najwyższą punktację otrzymuje rozwiązanie kompletne i poprawne. Niższy poziom obejmuje rozwiązania oparte na właściwej metodzie, lecz zawierające drobne błędy rachunkowe. Kolejny poziom zawiera rozwiązania świadczące o częściowym zrozumieniu problemu. Brak punktów oznacza, że przedstawiona metoda nie odpowiada treści zadania lub jest całkowicie przypadkowa.

Przykładowe poziomy oceniania:

Poziom 4 — pełne osiągnięcie celu

Uczeń rozpoznaje, że różnica wieku jest stała, poprawnie interpretuje warunek dotyczący dwukrotności wieku, stosuje właściwą metodę i otrzymuje prawidłowy wynik. Rozwiązanie jest jasno uzasadnione.

Poziom 3 — osiągnięcie celu z niewielkimi brakami

Uczeń stosuje poprawną metodę i uzyskuje właściwy wynik, ale uzasadnienie jest niepełne.

Poziom 2 — częściowe osiągnięcie celu

Uczeń zauważa, że różnica wieku wynosi 25 lat, albo poprawnie zapisuje sumę wieku, lecz nie potrafi połączyć obu informacji lub popełnia istotny błąd rachunkowy.

Poziom 1 — początkowe rozumienie

Uczeń podejmuje próbę rozwiązania, ale błędnie interpretuje relację „dwa razy tyle lat” albo wykorzystuje tylko jedną informację z zadania.

Poziom 0 — brak osiągnięcia celu

Uczeń nie przedstawia rozumowania związanego z zależnościami między wiekiem.

W przypadku analizowanej pracy można argumentować, że uczeń podjął próbę rozwiązania zadania, wykorzystał jedną z danych liczbowych i przeprowadził kontrolę wyniku. Z drugiej strony zasadnicza koncepcja rozwiązania jest błędna. W wielu schematach egzaminacyjnych taka praca otrzymałaby minimalną liczbę punktów przewidzianą za częściowe rozwiązanie.

Ocenianie kształtujące

Coraz większe znaczenie w edukacji zdobywa ocenianie kształtujące, którego głównym celem nie jest wystawienie stopnia, lecz wspieranie procesu uczenia się. W tym modelu nauczyciel udziela uczniowi informacji zwrotnej, wskazując mocne strony pracy, popełnione błędy oraz kierunki dalszego rozwoju.

W odniesieniu do omawianego rozwiązania nauczyciel mógłby podkreślić, że uczeń poprawnie wykorzystał informację o sumie wieku oraz wykazał się dbałością o sprawdzenie odpowiedzi. Jednocześnie należałoby zwrócić uwagę na błędne zrozumienie zdania „Alicja była wtedy o 25 lat młodsza” oraz na nieuzasadnione przyjęcie wieku 50 i 25 lat jako punktu wyjścia.

Kluczowym elementem takiej informacji zwrotnej byłaby wskazówka, aby przy kolejnych zadaniach rozpocząć rozwiązanie od oznaczenia niewiadomych i dokładnego opisanie sytuacji sprzed określonej liczby lat. Dzięki temu uczeń otrzymuje nie tylko ocenę swojej pracy, ale również konkretne narzędzia do poprawy własnych umiejętności.

Podsumowanie schematów oceniania wygenerowanych przez AI

Pokazana praca stanowi dobry przykład sytuacji, w której wybór schematu oceniania ma istotny wpływ na ostateczny rezultat. W modelu opartym wyłącznie na wyniku uczeń otrzymałby 0 punktów. W ocenianiu etapowym lub egzaminacyjnym mógłby uzyskać część punktów za poprawnie wykonane elementy rozwiązania. Z kolei ocena kompetencyjna i ocena kształtująca pozwalają dostrzec zarówno mocne strony pracy, jak i źródło popełnionego błędu. Z dydaktycznego punktu widzenia właśnie te ostatnie modele dostarczają najwięcej informacji o rzeczywistych umiejętnościach ucznia i najskuteczniej wspierają ucznia w dalszym uczeniu się matematyki.

Perspektywy wykorzystania sztucznej inteligencji w projektowaniu adaptacyjnych metod oceniania

Interesującym kierunkiem rozwoju metod oceniania jest zastosowanie AI jako systemu, który najpierw projektuje sposób oceniania konkretnego zadania, a następnie wykorzystuje go do oceny odpowiedzi.

AI jako narzędzie do tworzenia schematów oceniania

W przypadku ustalonego zadania sztuczna inteligencja może wygenerować różne schematy oceniania w zależności od tego, co chcemy mierzyć. To samo rozwiązanie można więc ocenić holistycznie, analitycznie albo za pomocą szczegółowego systemu punktowego. AI może przykładowo zaproponować kryteria dotyczące poprawności wyniku, poprawności poszczególnych etapów rozwiązania, zastosowanych metod, występujących błędów czy jakości uzasadnienia.

Jest to szczególnie interesujące w kontekście wcześniejszego problemu oceniania zero-jedynkowego. Odpowiedź końcowa może być poprawna, mimo że w toku rozwiązania pojawiły się błędy. Bardziej zaawansowany system AI mógłby nie ograniczać się do pytania: czy wynik jest poprawny? Ale analizować również: dlaczego wynik jest poprawny? Czy sposób dojścia do niego jest poprawny? Tym samym można przejść od automatycznego sprawdzania odpowiedzi do automatycznej analizy procesu rozwiązywania problemu.

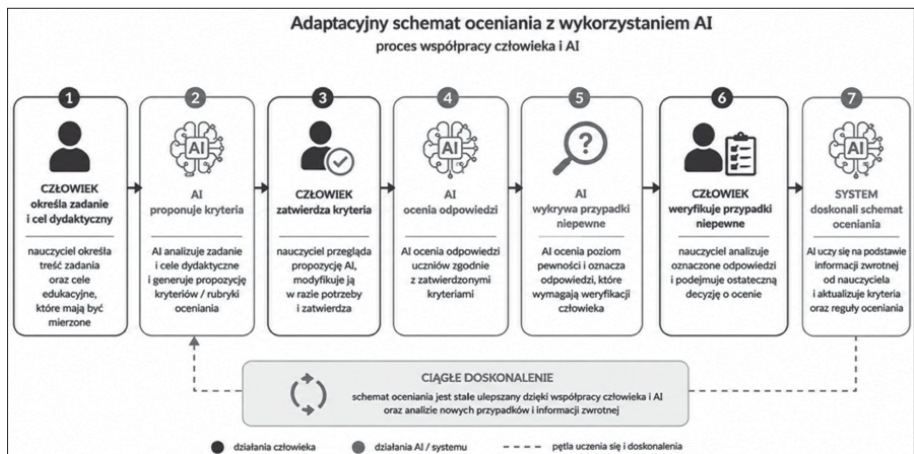
AI tworzy własne kryteria

W jakim kierunku rozwija się współczesne ocenianie? Obecnie często zakładamy schemat przedstawiony na rysunku 2.



Rysunek 2. Tradycyjny schemat oceniania z wykorzystaniem AI

W przyszłości może on wyglądać tak, jak przedstawiono na rysunku 3.



Rysunek 3. Adaptacyjny schemat oceniania z wykorzystaniem AI

W takim podejściu AI nie byłoby więc tylko automatycznym egzaminatorem. Pełniłoby funkcję projektanta i wykonawcy systemu oceniania, pozostawiając człowiekowi kontrolę nad tym, co właściwie powinno być mierzone.

Najnowsze prace naukowe mają zmierzać właśnie w podobnym kierunku. W EduMARS z 2026 r. zaproponowano podejście *Retrieval-Augmented Adaptive-Rubric Grading* (Zhao i in., 2026), w którym modele mogą dynamicznie tworzyć schematy oceny dostosowane do konkretnego przypadku. Z kolei system AutoSCORE (Wang i in., 2026) rozdziela proces na rozpoznawanie elementów odpowiedzi związanych z kryteriami oraz właściwe przyznawanie punktów. Autorzy wskazują, że takie uporządkowanie procesu zwiększa interpretowalność i zgodność z kryteriami.

AI tworzy kilka schematów oceniania jednego zadania

AI potrafi dla jednego zadania wygenerować kilka alternatywnych schematów oceniania, np.:

1. schemat wynikowy – czy uzyskano prawidłową odpowiedź;
2. schemat analityczny – ile etapów rozwiązania wykonano poprawnie;
3. schemat błędów – jakie błędy wystąpiły i jak poważne one były;
4. schemat kompetencyjny – jakie umiejętności uczeń rzeczywiście opanował;
5. schemat holistyczny – jaka jest ogólna jakość rozwiązania.

Następnie można porównać wyniki uzyskane za pomocą tych sposobów. Jeżeli wszystkie prowadzą do podobnej oceny, system może mieć dużą pewność swojej decyzji. Jeśli jednak jeden schemat daje np. 9/10, drugi 5/10, a trzeci 7/10, odpowiedź zostaje oznaczona jako przypadek wymagający sprawdzenia przez człowieka.

Takie podejście wydaje się bardziej wartościowe niż próba stworzenia jednego idealnego algorytmu oceniania.

AI mogłoby również uczyć się na swoich błędach w ocenianiu

Kolejnym etapem mogłoby być tworzenie systemu, który analizuje rozbieżności między własnymi ocenami a ocenami ekspertów. Jeżeli nauczyciel wielokrotnie poprawia ocenę AI w podobnym typie odpowiedzi, system może zidentyfikować brakujące kryterium i zaproponować modyfikację.

Przykładowo AI początkowo przyznaje pełny punkt za poprawny wynik. Nauczyciele zauważają jednak, że część uczniów dochodzi do niego poprzez dwa kompensujące się błędy. System mógłby wykryć powtarzalność takich przypadków i zaproponować nowe kryterium: Poprawny wynik końcowy nie jest wystarczający do przyznania pełnej liczby punktów, jeżeli wcześniejsze etapy rozwiązania zawierają błędy merytoryczne. Wtedy schemat oceniania nie byłby statyczny, lecz rozwijałby się wraz z pojawianiem się nowych typów odpowiedzi i błędów.

Podsumowanie

W kontekście prowadzonych rozważań istotnym kierunkiem badań wydaje się odejście od pytania ograniczającego rolę sztucznej inteligencji wyłącznie do automatycznego sprawdzania odpowiedzi: „Czy AI potrafi oceniać?”. Wraz z rozwojem dużych modeli językowych znacznie bardziej interesujące staje się pytanie: „Czy AI potrafi zaprojektować odpowiedni schemat oceniania określonego zadania, zastosować go w praktyce, określić poziom własnej niepewności, a następnie – na podstawie nowych przypadków oraz informacji zwrotnej – zaproponować jego udoskonalenie?”.

Szczególnie istotną funkcją przyszłych systemów mogłaby być zdolność do określania poziomu pewności własnej oceny. Zamiast traktowania każdej decyzji AI jako jednakowo wiarygodnej system mógłby rozpoznawać odpowiedzi nietypowe, niejednoznaczne lub takie, których nie można jednoznacznie zaklasyfikować na podstawie przyjętych kryteriów. Przypadki o wysokim poziomie pewności mogłyby być oceniane automatycznie, natomiast odpowiedzi o niskiej pewności byłyby kierowane do weryfikacji przez człowieka. W ten sposób rola nauczyciela lub eksperta nie zostałaby wyeliminowana, lecz przesunięta w stronę nadzoru nad najbardziej problematycznymi przypadkami.

Kolejnym etapem rozwoju mogłoby być wykorzystanie wyników tej weryfikacji do systematycznego doskonalenia schematu oceniania. Jeżeli człowiek wielokrotnie korygowałby decyzje AI dotyczące podobnego rodzaju błędów, system mógłby rozpoznać powtarzający się problem i zaproponować zmianę kryteriów. Schemat oceniania nie byłby zatem elementem statycznym, przygotowanym jednorazowo przed rozpoczęciem oceny, lecz strukturą rozwijaną wraz z pojawianiem się nowych odpowiedzi, strategii rozwiązania oraz wcześniej nieprzewidzianych błędów.

W tym ujęciu szczególnie interesującym przedmiotem przyszłych badań byłoby określenie, w jakim stopniu AI może samodzielnie projektować i modyfikować schematy oceniania, zachowując jednocześnie zgodność z ocenami ekspertów, przejrzystość podejmowanych decyzji oraz możliwość kontroli przez człowieka. Istotne byłoby również zbadanie stabilności takich ocen, ich odporności na nietypowe strategie rozwiązania oraz sposobów wyznaczania wiarygodnego poziomu niepewności modelu. W dłuższej perspektywie celem nie powinno być więc jedynie zastąpienie ręcznego przyznawania punktów automatyczną oceną, lecz stworzenie środowiska współpracy człowieka i AI, w którym system pomaga oceniać coraz trafniej, bardziej konsekwentnie i w sposób dostosowany do charakteru konkretnego zadania.

Tak rozumiane wykorzystanie sztucznej inteligencji pozwala spojrzeć na konkursy matematyczne „poza zwycięstwem” – nie tylko jako na mechanizm wyłaniania najlepszych uczestników i ustalania rankingów, lecz także jako na źródło wiedzy o sposobach matematycznego rozumowania. Analiza odpowiedzi może ujawniać różnorodne strategie rozwiązania, typowe i nietypowe błędy, częściowo poprawne ścieżki rozumowania oraz miejsca, w których uczniowie napotykały szczególne trudności. AI mogłaby wspierać identyfikowanie takich wzorców w skali trudnej do osiągnięcia wyłącznie poprzez pracę ekspertów.

Otrzymane w ten sposób wnioski stanowiłyby cenne narzędzie dla nauczycieli, którzy mogliby bezpośrednio przekładać je na praktykę dydaktyczną – wspierając uczniów w analizie zidentyfikowanych schematów, omówieniu najczęstszych pomyłek oraz prezentowaniu alternatywnych ścieżek rozumowania. Ogromną rolę mogliby tu odegrać sami organizatorzy konkursów, publikując materiały edukacyjne oparte na przeprowadzonych analizach i dając w ten sposób nauczycielom gotowe wsparcie w codziennej pracy. Przykładem inicjatywy wpisującej się w ten nurt jest publikacja „Kreatywność uczestników Pomorskiej Ligi Zadaniowej Zdolni z Pomorza. Uczniowie twórcami zadań” (Kreft, 2023). Dużym walorem tego opracowania jest prezentacja autorskich zadań uczniów wraz ze wzorcowymi rozwiązaniami i eksperckim komentarzem.

Literatura

- Hashemi, H., Eisner, J., Rosset, C., Van Durme, B., & Kedzie, C. (2024, August). LLM-rubric: A multidimensional, calibrated approach to automated evaluation of natural language texts. In *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 13806-13834).
- Jaksender, M., Zakrzewski R., Wypych-Stasiewicz A. (2016). Czy uczniowie powinni obawiać się nowego sposobu oceniania prac egzaminacyjnych? Wydział Chemii UJ, s. 159–164.
- Kreft A. (Red.) (2023), *Kreatywność uczestników Pomorskiej Ligi Zadaniowej Zdolni z Pomorza. Uczniowie twórcami zadań*, Pomorski Ośrodek Doskonalenia Nauczycieli w Słupsku, Słupsk.
- Pintal, D. (2022). *Ocenianie kształtujące. Od koncepcji do praktycznej realizacji w klasie zróżnicowanej*. Ośrodek Rozwoju Edukacji.
- Stańdo J. (2019). *Trajektorie i walidacja efektów uczenia się w kontekście Polskiej Ramy Kwalifikacji*, Wydawnictwo Politechniki Łódzkiej, Łódź.
- Stańdo J, Fechner Ż, Dąbrowicz-Tlałka A., Kujawska K. and Musielak M. (2025), *Exploring AI Chatbots for Learning Mathematics: Students' Perspectives on Accuracy and Educational Value*, 26th International Conference on Artificial Intelligence in Education.
- Szyling, G. (2011). *Nauczycielskie praktyki oceniania poza standardami*. Oficyna Wydawnicza Impuls.
- Wang, Y., Ding, Z., Wu, X., Sun, S., Liu, N., & Zhai, X. (2026, March). Autoscore: Enhancing automated scoring with multi-agent large language models via structured component recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 40, No. 48, pp. 40898-40906).
- Zhao, X., Chen, J., Yan, H., Fang, C., & Wei, X. (2026, July). EduMARS: Can Vision-Language Models Grade Like Teachers? Benchmarking Multimodal, Rubric-Based Assessment on Chinese K-12 Answers. In *Findings of the Association for Computational Linguistics: ACL 2026* (pp. 9561-9583).