

Czy sztuczna inteligencja zostawia ślady w tekście? Co nauczyciel powinien wiedzieć o „znakach wodnych” AI

Czy można rozpoznać, że wypracowanie ucznia napisała sztuczna inteligencja? Czy ChatGPT zostawia w tekście niewidzialny znak wodny? Czy charakterystyczny styl, długie myślniki albo idealnie uporządkowane akapity zdradzają użycie AI?

Wokół tych pytań narosło już sporo mitów. A sprawa jest ważna, szczególnie dla nauczycieli, którzy coraz częściej otrzymują prace przygotowane przy pomocy narzędzi generatywnej sztucznej inteligencji.

Spróbujmy więc uporządkować ten temat.

Znak wodny w tekście to nie logo na fotografii

Kiedy słyszymy „znak wodny”, najczęściej wyobrażamy sobie półprzezroczysty napis umieszczony na zdjęciu.

W przypadku tekstu generowanego przez sztuczną inteligencję sprawa wygląda zupełnie inaczej.

Duży model językowy tworzy tekst krok po kroku. W uproszczeniu przewiduje, jaki kolejny fragment, czyli token, najlepiej pasuje do tego, co zostało już napisane.

Technologia znakowania może delikatnie wpływać na prawdopodobieństwo wyboru określonych tokenów. W dłuższym tekście powstaje wtedy wzorzec statystyczny, którego człowiek nie widzi, ale odpowiednio przygotowany system może próbować go wykryć.

Tak działa na przykład rozwijana przez Google technologia **SynthID**. Google podkreśla jednocześnie, że nie jest ona uniwersalnym rozwiązaniem problemu rozpoznawania wszystkich treści generowanych przez AI. Najlepiej działa na dłuższych i bardziej zróżnicowanych wypowiedziach, a gruntowne przepisanie lub przetłumaczenie tekstu może osłabić możliwość wykrycia oznaczenia.

Warto dodać, że komponent SynthID odpowiadający za tekst został przez Google udostępniony jako otwarte narzędzie, z którego mogą korzystać także inne firmy. Nie zmienia to jednak najważniejszego wniosku: znak wodny działa tylko wtedy, gdy konkretne narzędzie AI faktycznie go stosuje, a odbiorca dysponuje odpowiednim detektorem. W praktyce szkolnej, przy pracy z różnymi modelami i bez dostępu do takich detektorów, nauczyciel nie ma z tego mechanizmu żadnego realnego pożytku.

Czy długi myślnik zdradza ChatGPT?

Nie.

Podobnie jak średnik, dwukropek, wypunktowania czy bardzo regularne akapity.

Modele językowe rzeczywiście mogą mieć charakterystyczne nawyki stylistyczne. Często tworzą uporządkowane wypowiedzi, lubią podsumowania, podobne konstrukcje zdań i przewidywalne schematy.

To jednak **cechy stylu, a nie dowód autorstwa**.

Uczeń również może pisać bardzo poprawnie. Człowiek może używać średników. Nauczyciel może lubić rozbudowane wypunktowania.

I odwrotnie – AI można poprosić o tekst potoczny, mniej uporządkowany albo napisany określonym stylem.

Dlatego nie powinniśmy wyciągać wniosku:

„Ten tekst wygląda jak ChatGPT, więc napisała go sztuczna inteligencja”.

To za mało.

A może wystarczy detektor AI?

Tu potrzebna jest jeszcze większa ostrożność.

W internecie znajdziemy serwisy, które po wklejeniu tekstu pokazują na przykład:

„Prawdopodobieństwo użycia AI: 94%”.

Taki wynik może wyglądać bardzo przekonująco.

Nie oznacza jednak automatycznie, że mamy 94-procentową pewność, iż uczeń wykorzystał ChatGPT.

Nawet Google, opisując SynthID, zwraca uwagę, że klasyfikatory służące do rozpoznawania tekstów AI nie zawsze działają równie dobrze dla różnych rodzajów treści i mogą prowadzić do błędnego oznaczenia tekstu.

Problem nie jest wyłącznie teoretyczny. Głośny w środowisku akademickim test kilku popularnych detektorów wykazał, że jeden z nich fałszywie oznaczał jako „AI” niemal połowę tekstów napisanych w rzeczywistości przez ludzi. Zespół badaczy ze Stanford University sprawdził z kolei, jak detektory radzą sobie z esejami uczniów, dla których angielski nie jest językiem ojczystym – okazało się, że ponad połowę takich prac oznaczono jako wygenerowane przez AI, podczas gdy teksty rodzimych użytkowników języka klasyfikowano niemal bezbłędnie.

To ważna wskazówka również dla polskiej szkoły, zwłaszcza tam, gdzie uczą się dzieci dwujęzyczne, uczniowie z doświadczeniem migracyjnym albo osoby ze specyficznymi trudnościami w nauce. Nietypowy, prostszy albo bardzo skonwencjonalizowany styl pisanie wcale nie musi oznaczać użycia AI – a detektor może to zinterpretować dokładnie odwrotnie.

Dlatego wynik detektora może być **sygnałem do dalszego sprawdzenia**, ale nie powinien sam w sobie przesądzać o autorstwie pracy ucznia.

Zmieńmy pytanie

Być może zamiast pytać:

„Jak udowodnić, że uczeń użył AI?”

warto częściej zapytać:

„Czy uczeń rozumie pracę, którą oddał?”

To niewielka zmiana pytania, ale ogromna zmiana podejścia.

Wyobraźmy sobie, że uczeń przynosi świetnie napisane wypracowanie.

Zamiast rozpoczynać śledztwo z użyciem trzech detektorów AI, możemy powiedzieć:

- *„Wyjaśnij własnymi słowami ten fragment”.*
- *„Dlaczego wybrałeś właśnie ten argument?”*
- *„Skąd pochodzi ta informacja?”*
- *„Podaj inne źródło potwierdzające tę tezę”.*
- *„Zmieńmy warunek zadania. Jak teraz zmieniłaby się Twoja odpowiedź?”*
- *„Napisz tutaj, na lekcji, krótkie podsumowanie swojej pracy”.*

Bardzo szybko dowiemy się, czy rozmawiamy z **autorem pracy**, czy tylko z osobą, która dostarczyła gotowy tekst.

AI nie musi być zakazane. Musi być świadomie używane

Nie chodzi przecież o to, aby uczeń nigdy nie korzystał ze sztucznej inteligencji.

- Może wykorzystać AI do zebrania pomysłów.
- Może poprosić o wyjaśnienie trudnego pojęcia.
- Może poszukać kontrargumentów.
- Może poprosić o informację zwrotną do własnego tekstu.
- Może wykorzystać model do poprawienia języka.

Problem zaczyna się wtedy, gdy znika aktywność ucznia:

prompt → kopiuj → wklej → oddaj.

Wtedy nie bardzo wiemy, czego nauczył się uczeń.

Prosta zasada: AI → SPRAWDŹ → PRZEREDAGUJ → DODAJ SIEBIE → UJAWNIJ

Taką zasadę warto zaproponować również uczniom.

SPRAWDŹ. AI może się pomylić. Sprawdź nazwiska, daty, liczby, cytaty i źródła. Jeżeli model podaje publikację naukową, sprawdź, czy rzeczywiście istnieje i czy mówi to, co twierdzi AI.

PRZEREDAGUJ. Nie chodzi o zamianę kilku słów na synonimy. Najpierw zrozum treść. Potem przedstaw ją własnym językiem.

DODAJ SIEBIE. Dodaj własny przykład, argument, doświadczenie, interpretację albo wniosek. To właśnie tutaj zaczyna być widoczna praca człowieka.

Dodałbym jeszcze jeden element:

UJAWNIJ.

Jeżeli zasady wykonania zadania dopuszczają korzystanie z AI, uczeń może po prostu napisać:

„Przy przygotowaniu pracy wykorzystałem AI do zebrania pomysłów i sprawdzenia struktury tekstu. Ostateczną wersję opracowałem samodzielnie”.

To uczy odpowiedzialności znacznie lepiej niż zabawa w policjanta i złodzieja.

Nie uczmy sposobów „oszukiwania detektorów”

W sieci można znaleźć mnóstwo porad:

- „usuń długie myślniki”,
- „dodaj kilka błędów”,
- „poproś AI, żeby pisała bardziej po ludzku”,
- „przepuść tekst przez humanizer”,
- „parafrazuj, aż detektor pokaże zero”.

To ślepa uliczka.

Uczeń nie uczy się wtedy świadomego korzystania ze sztucznej inteligencji. Uczy się **ukrywania faktu jej wykorzystania**.

A przecież naszym celem edukacyjnym powinno być coś zupełnie innego.

Chcemy, żeby uczeń potrafił wykorzystać AI, ale jednocześnie umiał ocenić jej odpowiedź, wykryć błąd, sprawdzić źródło, zakwestionować wynik i podjąć własną decyzję.

Dochodzi jeszcze AI Act

Temat oznaczania treści generowanych przez sztuczną inteligencję przestał być wyłącznie technologiczną ciekawostką.

Od **2 sierpnia 2026 r.** stosowane są nowe obowiązki dotyczące przejrzystości wynikające z art. 50 unijnego AI Act. Między innymi dostawcy określonych systemów generatywnych mają zapewniać oznaczanie treści generowanych lub manipulowanych przez AI w formacie nadającym się do odczytu maszynowego. Komisja Europejska przyjęła w lipcu 2026 roku szczegółowe wytyczne wyjaśniające, jak stosować te przepisy w praktyce.

Nie oznacza to jednak, że od tej daty **każdy tekst napisany przy pomocy AI musi mieć widoczny napis „wygenerowano przez sztuczną inteligencję”**.

Przepisy są bardziej szczegółowe. Szczególne obowiązki informacyjne dotyczą między innymi tekstów publikowanych w celu poinformowania społeczeństwa o sprawach interesu publicznego, gdy nad treścią nie czuwa człowiek ani nie odpowiada za nią redakcja. Dla dostawców systemów, które wygenerowano jeszcze przed 2 sierpnia 2026 r., obowiązek technicznego znakowania treści odroczone do 2 grudnia 2026 r.

Dla szkoły płynie z tego szerszy wniosek.

Transparentność korzystania ze sztucznej inteligencji będzie coraz ważniejsza.

Co zatem powinien zrobić nauczyciel?

Nie zaczynałbym od instalowania kolejnego detektora AI.

Zacząłbym od ustalenia zasad.

Uczeń powinien wiedzieć przed wykonaniem zadania:

- czy może korzystać z AI;
- do czego może jej użyć;
- czego AI nie może wykonać za niego;
- czy powinien wskazać sposób wykorzystania AI;
- za które elementy pracy odpowiada sam;
- czy będzie musiał wyjaśnić lub obronić swoją pracę.

Jeszcze lepszym rozwiązaniem jest projektowanie takich zadań, w których **samo wygenerowanie odpowiedzi nie wystarcza**.

Poprośmy o uzasadnienie.

O porównanie dwóch rozwiązań.

O odniesienie do doświadczenia z lekcji.

O znalezienie błędu w odpowiedzi AI.

O poprawienie odpowiedzi modelu.

O wskazanie źródeł.

O wykonanie kolejnego etapu zadania na lekcji.

Wtedy AI może pomagać uczniowi, ale nie przejmuje całego procesu uczenia się.

I chyba właśnie tutaj leży sedno

Prawdopodobnie jeszcze długo będziemy obserwować wyścig między modelami generującymi tekst a systemami próbującymi ustalić jego pochodzenie.

Powstaną lepsze znaki wodne.

Powstaną lepsze detektory.

I zapewne powstaną również lepsze sposoby modyfikowania wygenerowanych treści.

Szkoła nie powinna jednak budować swojej strategii na nadziei, że zawsze będzie o krok przed uczniem.

Mamy lepsze narzędzie niż detektor. Jest nim dobrze zaprojektowany proces uczenia się.

Jeżeli uczeń potrafi wyjaśnić swoją pracę, obronić argumenty, wskazać źródła, znaleźć błąd AI, poprawić go i wykorzystać zdobytą wiedzę w nowej sytuacji, otrzymujemy znacznie więcej informacji o jego rzeczywistych kompetencjach niż z komunikatu:

„AI detector: 87%”.

Dlatego pytanie, które będziemy sobie coraz częściej zadawać w szkole, nie powinno brzmieć:

„Czy to napisała sztuczna inteligencja?”

Lepiej zapytać:

„Co w tej pracy zrobiła sztuczna inteligencja, a czego nauczył się i za co odpowiada uczeń?”

To drugie pytanie jest trudniejsze.

Ale właśnie dlatego jest znacznie bardziej edukacyjne.

W skrócie – cztery rzeczy do zapamiętania:

1. Styl to nie dowód. Długie myślniki, wypunktowania czy uporządkowane akapity nie potwierdzają użycia AI.

2. Detektorom nie ufaj bezkrytycznie. Mylą się częściej wobec tekstów uczniów dwujęzycznych, migrantów i osób ze specyficznymi trudnościami w nauce.

3. Pytaj o proces, nie tylko o efekt. Krótka rozmowa o pracy ucznia mówi więcej niż wynik jakiegokolwiek skanera.

4. Ucz zasady „SPRAWDŹ → PRZEREDAGUJ → DODAJ SIEBIE → UJAWNIJ”. To ona, a nie zakaz, uczy odpowiedzialnego korzystania z AI.

Dbajmy o uczniów, ale nie zapominajmy o sobie.

Mariusz Walak