1.3 Odpowiedzialne wykorzystanie SI

Najważniejsze przesłanie

- **Odpowiedzialność to zadanie projektowe, a nie kontrolne.** Wykrywanie SI nie działa poprawnie.
- **A wykrywanie SI jest niesprawiedliwe, gdy zawodzi.** Detektory błędnie oznaczają pracę człowieka jako napisaną przez SI i przeoczają prace napisane przez SI. Władze oświatowe we Francji i w Holandii odradzają ich stosowanie.
- **Pewne siebie zmyślanie i odziedziczone uprzedzenia wynikają z tego, jak te systemy są budowane i oceniane.** Konsekwencją powinno być: uczyć o nich, zamiast czekać na naprawę.
- **Zasada działania brzmi: najpierw myśl, potem SI.** Zobowiązanie ludzi do własnej odpowiedzi bije wyjaśnianie działania maszyny, a wbudowanie ograniczeń w oprogramowanie usuwa szkody, jakie wyrządza nieograniczony dostęp.
- **Same polecenia (prompty) nie wystarczą.** Chatbot mający pełnić rolę korepetytora ma tendencję do powracania do podawania gotowych odpowiedzi, co sprawia, że niezawodne ograniczenia stają się kwestią zakupu narzędzia, a nie umiejętności formułowania poleceń.
- **Wzmacnianie różnic jest domyślne; wyrównywanie trzeba zbudować** z dostępu, projektu i wsparcia razem. Najrzadszym, ale najważniejszym z tych trzech elementów jest przeszkolony nauczyciel.
- **Włączanie (inkluzja) to najwyraźniejsza korzyść** i to właśnie tu sami nauczyciele widzą najmocniejsze argumenty.

> *Jeśli zapamiętasz z tej strony tylko jedno zdanie:* Tamy pękają. Naucz pływania.

Dziennik lektur, drugie spojrzenie

Wróćmy do naszego ucznia z dziewiątej klasy i jego podejrzanie dopracowanego dziennika lektur. Ten odruch jest niemal powszechny i jest to odruch dobrego nauczyciela: sprawdzić, czy użył SI. Zbudować lepszą tamę.

Ten rozdział opowiada o tym, dlaczego ten odruch, sam w sobie, donikąd nie prowadzi. Dojdziemy do tego, co działa zamiast tego.

Zacznijmy od niewygodnej części. **Nauczyciele nie potrafią wiarygodnie odróżnić tekstu SI od tekstu ucznia.** W dwóch badaniach niemieckiego zespołu 89 nauczycieli w trakcie przygotowania do zawodu i 200 doświadczonych nauczycieli otrzymało mieszankę wypracowań uczniów i wypracowań napisanych przez chatbota i poproszono ich o ich posortowanie. Żadna z grup nie potrafiła zrobić tego wiarygodnie, ale obie grupy były pewne, że potrafią. Doświadczenie pomagało tylko w niewielkim stopniu: doświadczeni nauczyciele wciąż przeoczyli większość tekstów SI. Na jednym brytyjskim uniwersytecie badacze przesłali odpowiedzi wygenerowane przez chatbota do prawdziwych, nienadzorowanych egzaminów online zwykłymi kanałami, bez informowania osób sprawdzających.

Głosy z praktyki
„Możesz z tego korzystać, ale musisz umieć wyjaśnić wynik własnymi słowami.”

Detektor, który zawodzi wobec niewłaściwych uczniów

Badacze przetestowali siedem najczęściej używanych detektorów SI na esejach TOEFL napisanych przez osoby uczące się angielskiego jako języka obcego, zebranych z chińskiego forum internetowego. Obok nich umieścili eseje amerykańskich ósmoklasistów mówiących po angielsku jako języku ojczystym. Prace rodzimych użytkowników języka były klasyfikowane poprawnie niemal za każdym razem. Spośród esejów obcojęzycznych ponad sześć na dziesięć zostało błędnie oznaczonych jako napisane przez maszynę.

Mechanizm nie jest tajemnicą. Wiele z tych narzędzi nie wykrywa pracy dużych modeli językowych. Narzędzia te mierzą *przewidywalność*. Tekst używający powszechnych słów w oczekiwanym porządku jest oceniany jako sztuczny. Osoby piszące w drugim języku, dysponujące mniejszym słownictwem i bardziej sztywnymi konstrukcjami zdań, mogą wytwarzać dokładnie taki sygnał. Tymczasem detektory wytrenowane pod kątem jednego konkretnego testu, a także badania w innych językach, wykazują mniejsze różnice. Jak dotąd na rynku nie ma detektorów na tyle wiarygodnych, by na ich podstawie oskarżać kogokolwiek o celowe oszustwo.

Błędy narzędzia systematycznie spadają na uczniów, którzy i tak dźwigają już najwięcej. Instytucje policzyły to dokładnie. Uniwersytet Vanderbilt zwrócił uwagę, że nawet gdyby takie narzędzie myliło się tylko w 1% przypadków, oznaczałoby to wciąż około **750 prac rocznie błędnie oznaczonych** spośród 75 000 przesłanych prac. W związku z tym je wyłączono. Przewodnik opracowany przez badaczy dla niemieckiego federalnego ministerstwa edukacji jest jednoznaczny: technologia jest *„niewystarczająco wiarygodna, by dostarczyć prawnie pewny dowód”*. Niezależny europejski test czternastu detektorów doprowadził do tego samego werdyktu:

„ani dokładne, ani wiarygodne”.

Znaczna część Europy przestała już o tym dyskutować. Francuskie ministerstwo edukacji i francuski organ ochrony danych odradzają szkołom stosowanie detektorów, a holenderska agencja ds. technologii w edukacji Kennisnet uznaje je za wysoce zawodne. W szkolnictwie wyższym niektóre uczelnie, takie jak ZHAW w Zurychu, zakazują wykorzystywania ich do oceniania. W Anglii komisje egzaminacyjne dopuszczają oprogramowanie do wykrywania jedynie jako jeden z elementów osądu, który należy do nauczyciela, ponieważ to on *„najlepiej zna swoich uczniów”*. Warto też zapamiętać kwestię ochrony danych: **wklejenie pracy ucznia do narzędzia online oznacza przetwarzanie jego danych osobowych.** Wymaga to podstawy prawnej i umowy z dostawcą. → 1.4 Zgodne wykorzystanie SI

Wygląda na to, że panika związana ze ściąganiem jest cieńsza niż szum wokół niej. Badacze przepytali uczniów trzech amerykańskich szkół średnich przed powstaniem ChatGPT i ponownie po nim. Ściąganie pozostało w dużej mierze na tym samym poziomie. Kolejne badanie ankietowe, przeprowadzone rok później wśród ponad 4000 uczniów sześciu innych amerykańskich szkół średnich, wykazało to samo.

Prawdziwe pytanie nigdy nie brzmiało: „Jak go złapać?”. Brzmi ono: co właściwie mierzy zadanie domowe, skoro jego produkt można wygenerować na żądanie?

Głosy z praktyki

„Na początku uważałem to za ściąganie. Teraz widzę w tym sygnał, że trzeba zmienić samo zadanie.”

Dwie rzeczy, których nie da się po prostu naprawić, więc trzeba ich nauczać

Pewne siebie zmyślanie jest wbudowane w system. Te systemy generują prawdopodobne kontynuacje tekstu, a prawdopodobne nie znaczy prawdziwe. Przeglądy badań są zgodne, że żadna technika nie wyeliminowała dotąd tego problemu, choć można go ograniczyć i częściowo wykrywać. Powodem jest to, że standardowe testy służące do oceny dużych modeli językowych nagradzają pewne siebie zgadywanie bardziej niż przyznanie się do niepewności, więc dopóki modele są oceniane w ten sposób, utrzyma się pewien odsetek płynnych, dobrze sformułowanych nieprawd.

Skrzywienia w materiale treningowym również są wbudowane. Co więcej, nie maleją wraz z doskonaleniem modeli. Europejskie badanie stereotypów płciowych w różnych językach wykazało, że większe modele, nawet po dodatkowym treningu mającym uczynić je bezpieczniejszymi i sprawiedliwszymi, czasem stereotypizują bardziej, a nie mniej. Jeśli jakiejś cechy nie da się zaprojektować na nowo, staje się ona materiałem programowym. I można jej nauczać: w jednym

fińskim programie odsetek uczniów potrafiących wyjaśnić, skąd wzięło się uprzedzenie generatora obrazów, wzrósł z 7% do 44%. Konsekwencją dla nauczania musi być trenowanie analizy błędów zamiast ich unikania. Innymi słowy: błędy maszyny stają się lekcją.

Głosy z praktyki

„Niech uczący się sprawdzi rozwiązanie chatbota tak, jakby był nauczycielem. Szukając błędów, w końcu uczy się metody.”

Co działa: projektowanie, nie kontrola

Dowody wskazują na jedną prostą zasadę, łatwą do zapamiętania: najpierw myśl, potem SI.

- **Spraw, by człowiek zobowiązał się jako pierwszy.** W eksperymencie z udziałem około 200 dorosłych osoby, które musiały podjąć własną decyzję, *zanim* zobaczyły sugestię maszyny, rzadziej przyjmowały jej błędne sugestie niż osoby, którym pokazano krótkie wyjaśnienie toku rozumowania maszyny. Wyjaśnianie działania SI nie rozwiązało problemu. Pomogła metoda wymuszenia własnego osądu. Wymuszenie osądu oznacza zmuszenie ludzi do myślenia. Ponieważ myślenie jest męczące, uczestnicy lubili te metody najmniej, a najbardziej pomagały one tym, którzy i tak już lubią wysiłek poznawczy. To zjawisko znane jako efekt Mateusza: temu, kto ma, będzie dodane.
- **Wbuduj ograniczenia w system.** Chatbot z rozdziału 1.2, który nie podawał odpowiedzi, lecz zadawał pytania, to ten sam pomysł zrealizowany w oprogramowaniu: uczniowie, którzy z niego korzystali, nic nie stracili na egzaminie, podczas gdy ci z nieograniczoną wersją stracili bardzo dużo.
- **Uczyń proces widocznym, zamiast tropić produkt.** Szkice, notatki robocze, krótka rozmowa o pracy. Pokazują one, co uczeń potrafi, bez konieczności udowadniania, co dokładnie zrobił.
- **Dania pilotuje tę zasadę:** w uczestniczących szkołach uczniowie mogą korzystać z SI podczas przygotowań do ustnego egzaminu z angielskiego, a egzamin pisemny obejmuje część odręczną, bez pomocy.

Warto znać dwa ograniczenia, zanim się na tym oprzesz.

Pierwsze: popularna rada, by nauczyciele po prostu polecili chatbotowi pełnić rolę tutora sokratejskiego, może nie wytrzymać próby dłuższej rozmowy. Wstępne dowody z niezrecenzowanego jeszcze badania z symulowanymi uczniami (październik 2026) wskazują, że pod uporczywą presją tacy tutorzy wracają do podawania gotowych odpowiedzi. Niezawodne ograniczenia trzeba wbudować w system, co czyni z nich decyzję zakupową, a nie umiejętność formułowania poleceń.

Drugie: sprawienie, by SI wyjaśniała swój tok rozumowania, brzmi jak oczywiste zabezpieczenie, ale wyjaśnienia niezawodnie nie zapobiegają nadmiernemu zaufaniu. W jednym badaniu nie

zmieniły one wcale skłonności ludzi do akceptowania *błędnych* porad, a wyjaśnienia zbyt techniczne albo zbyt uproszczone mogą wręcz zwiększać nieuzasadnioną pewność. Wyjaśnienie jest jedynie warunkiem uczenia się.

Postawa: dwa tryby zawodności, nie jeden

Obawiamy się, że nauczyciele pokładają w SI zbyt duże zaufanie. Powinniśmy jednak równie mocno martwić się sytuacją odwrotną. Nawet doświadczone osoby mogą dać się złapać na nadmierne zaufanie. W niewielkim pilotażu w amerykańskich szkołach czterech na dziesięciu ankietowanych nauczycieli uznało informację zwrotną narzędzia SI do oceniania za nieprzydatną, przeważnie niejasną, a nawet błędną. Badania nad błędem automatyzacji (automation bias) pokazują, że ludzie mają tendencję do przyjmowania sugestii maszyny nawet wtedy, gdy mają powody, by w nią wątpić. Dostrzeżenie błędu i jego poprawienie to dwa różne kroki.

Przeciwnym biegunem jest ukrywanie. Nauczyciele ukrywają własne korzystanie z SI przed kolegami i uczniami, zabraniając im robić to, co sami robią: 77% korzysta z SI prywatnie, podczas gdy 57% twierdzi, że uczniowie nigdy nie powinni jej używać do generowania pomysłów. Tymczasem badanie zdolnych uczniów klas 6–8 wykazało, że od jednej trzeciej do dwóch piątych z nich już teraz twierdzi, że SI wie więcej niż ich nauczyciele. Zawód, który ukrywa swoje praktyki, nie może dawać przykładu odpowiedzialnego postępowania, a przejrzystość to nie tylko kurtuazja. To mechanizm, dzięki któremu działa nauczanie przez przykład.

Czy SI wyrównuje szanse, czy je pogłębia?

W zadaniach zawodowych SI wyrównuje różnice. W jednym eksperymencie osoby z wyższym wykształceniem radziły sobie znacząco lepiej niż osoby z niższym wykształceniem. Z asystentem SI zniknęły około trzech czwartych tej różnicy. Gdy asystenta ponownie odebrano, znaczna część różnicy powróciła. To, co SI zamknęła, to luka w wynikach, a nie luka w umiejętnościach.

W zadaniach związanych z uczeniem się dzieje się odwrotnie. Jedno badanie nie wykazało żadnej przeciętnej korzyści, za to ujawniło rosnący dystans między uczniami, którzy już dużo wiedzieli, a tymi, którzy wiedzieli mało. Inne badanie przyjrzało się uczniom piszącym w języku obcym: słabsi autorzy otrzymywali od SI więcej poprawek i skutecznie wykorzystywali mniejszą ich część, a wielu uznało zalew informacji zwrotnej za zniechęcający. Informacja zwrotna bez umiejętności jej posortowania nie pomaga – ona przytłacza.

Ta różnica istnieje jeszcze zanim ktokolwiek dotknie narzędzia. W Niemczech 80% uczniów z najbardziej uprzywilejowanych domów postrzega SI jako szansę, wobec 55% z domów najmniej uprzywilejowanych. W Wielkiej Brytanii podział przebiega raczej przez pokój nauczycielski niż przez szafkę z urządzeniami: 45% nauczycieli w szkołach prywatnych przeszło formalne szkolenie z korzystania z SI, wobec 21% w szkołach publicznych. W obrębie sektora publicznego zamożniejsze szkoły ponownie wyprzedzają biedniejsze. Badanie z 2026 roku obejmujące ponad 25 000 dzieci z 17 krajów europejskich pokazuje ten sam społeczny podział w tym, jak często i jak różnorodnie korzystają z SI. Zdanie jednego z komentatorów ujmuje cały problem lepiej niż jakakolwiek liczba:

„bogaci mają dostęp zarówno do technologii, jak i do ludzi, którzy pomagają im z niej korzystać, podczas gdy biedni mają dostęp jedynie do technologii”.

Wąskim gardłem jest więc to, czy dany uczeń ma nauczyciela, który przeszedł szkolenie. SI kompensuje różnice, gdy spełnione są jednocześnie trzy warunki: pewny dostęp, narzędzie zaprojektowane do uczenia się, a nie do udzielania odpowiedzi, oraz nauczyciel wspierający jego wykorzystanie. Usuń choć jeden z tych elementów, a SI zacznie wzmacniać różnice zamiast je wyrównywać. Ponieważ wszystkie trzy elementy zwykle występują razem w tych samych szkołach i razem są nieobecne w tych samych szkołach, wzmacnianie różnic jest wariantem domyślnym, a wyrównywanie trzeba samodzielnie zbudować.

Revision #14
Created 2026-09-01 09:14:31 UTC by Alexander
Updated 2026-10-20 15:40:04 UTC by Alexander