

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.** Istnieje ryzyko błędnego uznania odpowiedzi za wygenerowaną przez SI lub odwrotnie. Kilka europejskich krajów już zrezygnowało z tych narzędzi.
- **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; wbudowanie ograniczeń w oprogramowanie bije jego zakazywanie.
- **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 – a najrzadszym 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 podejrzenie 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 niemieckim badaniu 289 nauczycieli o różnym doświadczeniu otrzymało mieszankę tekstów napisanych przez uczniów i przez chatbota i poproszono ich o posortowanie. Żadna z grup sobie z tym nie poradziła. Ale obie grupy były pewne, że potrafią. Doświadczenie niemal nie miało znaczenia. W większej skali jest jeszcze gorzej. Na jednym uniwersytecie badacze przesłali odpowiedzi wygenerowane przez chatbota do prawdziwego systemu egzaminacyjnego zwykłymi kanałami, bez informowania osób sprawdzających. 94% nigdy nie wzbudziło wątpliwości. Co więcej, średnio oceniano je wyżej niż prace prawdziwych studentów. Jaka więc jest konsekwencja? Sięgamy po maszynę! I w tym momencie odkrycie przestaje być tylko niezręczne, a staje się problemem etycznym!

Detektor, który zawodzi wobec niew?a?ciwych uczniów

Badacze przetestowali siedem najczęściej używanych detektorów SI na esejach TOEFL napisanych przez uczniów uczących się angielskiego jako języka obcego. 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. Te narzędzia nie wykrywają pracy dużych modeli językowych, wykrywają przewidywalność. Tekst używający powszechnych słów w oczekiwanym porządku jest oceniany jako sztuczny. Każdy, kto pisze w drugim języku, dysponując mniejszym słownictwem i bezpieczniejszymi konstrukcjami zdań, wytwarza dokładnie taki sygnał. 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 niesłusznie oskarżonych studentów rocznie spośród 75 000 przesłanych prac. W związku z tym je wyłączono. Wytyczne wydane przez niemieckie federalne ministerstwo edukacji są jednoznaczne: technologia jest „niewystarczająco wiarygodna, by dostarczyć prawnie pewny dowód”.

Znaczna część Europy przestała już o tym dyskutować. Francja, Holandia i Szwajcaria odradzają stosowanie detektorów albo je zakazują. Brytyjski regulator egzaminów świadomie wybrał ludzki osąd zamiast oprogramowania. Holenderski argument idzie o krok dalej i warto go zapamiętać: wklejenie pracy ucznia do detektora samo w sobie stanowi przetwarzanie jego danych osobowych, a robienie tego bez zgody łamie prawo ochrony danych. → [1.4 Zgodne wykorzystanie SI](#)

Wygląda na to, że panika związana z ściąganiem jest cieńsza niż szum wokół niej. Badacze przepytali uczniów trzech szkół średnich przed powstaniem ChatGPT i ponownie po nim. Ściąganie pozostało w dużej mierze na tym samym poziomie. Większość uczniów odrzucała pomysł, by chatbot po prostu wykonał za nich pracę, uważając jednocześnie za uczciwe użycie go do rozpoczęcia pracy albo do uzyskania wyjaśnienia.

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

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

Pewne siebie zmyślanie jest wbudowane w system. Te systemy generują prawdopodobną kontynuację tekstu, co nie jest tym samym, co prawdę. Badacze formalnie wykazali już, skąd się to bierze: standardowe testy służące do oceny dużych modeli językowych nagradzają pewne siebie zgadywanie bardziej niż przyznanie się do niepewności, więc dopóki są oceniane w ten sposób, utrzymuje się pewien odsetek płynnych, dobrze sformułowanych nieprawd. To cecha sposobu, w jaki są budowane i oceniane, a nie błąd czekający na łatkę.

Skrzywienia w materiale treningowym również są wbudowane. Co więcej, nie maleją wraz z rozwojem technologii. 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 rozpoznać uprzedzenie w wyniku działania systemu SI wzrósł z 7% do 44%. Konsekwencją dla nauczania musi być trenowanie analizy błędów zamiast ich unikania. Innymi słowy: błędy maszyny stają się lekcją.

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 serii eksperymentów osoby, które musiały zapisać własną odpowiedź, zanim zobaczyły sugestie maszyny, wychwytywały znacznie więcej jej błędów niż osoby, którym pokazano szczegółowe wyjaśnienia toku rozumowania maszyny. Wyjaśnianie działania SI nie pomogło rozwiązać tego problemu, ale wymuszenie własnego osądu – tak. Z jednym uczciwym zastrzeżeniem: wersje, które działały najlepiej, były tymi, które uczestnicy lubili najmniej, i pomagały najbardziej tym, którzy i tak już lubią wysiłek poznawczy. Jest to więc zjawisko znane też jako efekt Mateusza: temu, kto ma, będzie dodane. To oni korzystają z tego podejścia najbardziej.
- **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 odrębną, bez pomocy.

Warto znać dwa ograniczenia, zanim się na tym oprzesz. Pierwsze: popularna rada, by nauczyciele po prostu polecieli chatbotowi pełnić rolę tutora sokratejskiego, nie wytrzymuje próby dłuższej rozmowy. Przynajmniej obecnie system ma tendencję do powracania 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 warunkiem, nie rozwiązaniem.

Postawa: dwa tryby zawodno?ci, 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 jednym badaniu nauczyciele sprawdzali oceny wygenerowane przez SI. Uznali 42% informacji zwrotnej maszyny za niejasną lub błędną, ale zmienili tylko 9% z nich. Dostrzeżenie błędu i jego poprawienie to dwa różne procesy.

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 pog??bia?

W zadaniach zawodowych SI wyrównuje różnice. W jednym eksperymencie osoby z wyższym wykształceniem radziły sobie znacząco lepiej niż osoby z niższym wykształceniem. Gdy każda z nich otrzymała asystenta SI, te różnice niemal całkowicie zniknęły. Gdy następnie odebrano asystenta, znaczna ich część 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 dzieci z 20 krajów europejskich pokazuje ten sam społeczny podział w tym, jak często i jak różnorodnie korzystają z SI. Jedno zdanie ujmuje cały problem lepiej niż jakakolwiek liczba: „bogaci mają dostęp do technologii i do ludzi, którzy pomagają im z niej korzystać, podczas gdy biedni mają dostęp jedynie do technologii”. To najmocniejszy dostępny argument za tym, że nauczyciele mają znaczenie dla zamknięcia tej luki.

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. Usunąć 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 #11

Created 2026-09-01 09:14:31 UTC by Alexander

Updated 2026-09-02 09:50:25 UTC by Alexander