

„Antyplagiat” z nową funkcją wykrywania treści pisanych przez AI

27 lutego 2024

Łatwiej będzie wykryć treści pisane przez sztuczną inteligencję w pracach dyplomowych. Taki moduł został właśnie dodany do Jednolitego Systemu Antyplagiatowego, który jest obligatoryjnie wykorzystywany do sprawdzania samodzielności w pisaniu takich prac. Pomoc sztucznej inteligencji w tym procesie, zdaniem ekspertów, nie jest niczym złym, ale już generowanie treści, pod którą autor tylko się podpisuje, świat nauki traktuje nagannie. Nie ma jak na razie żadnych regulacji, które normowałyby te kwestie.

„Nie jesteśmy w stanie dokładnie oszacować skali użycia sztucznej inteligencji w procesie tworzenia pisemnych prac dyplomowych, natomiast to zjawisko z pewnością istnieje i wzrasta jego siła. Wnioskujemy to na podstawie kontaktów ze środowiskiem i zgłoszeń, które otrzymujemy od przedstawicieli uczelni, uczestnicząc w różnego typu konferencjach, konsultacjach, również przez nasz dział wsparcia użytkowników. Generalnie ciężko zakładać, że cała praca jest możliwa do wygenerowania przez studenta za pomocą generatywnych modeli sztucznej inteligencji, raczej są to fragmenty” – ocenia w rozmowie z agencją Newseria Marek Michajłowicz, zastępca dyrektora Ośrodka Przetwarzania Informacji – Państwowego Instytutu Badawczego.

Zainteresowanie studentów wykorzystaniem generatywnej sztucznej inteligencji w pisaniu prac dyplomowych jest całkiem spore. Z ubiegłorocznego raportu „Technologia okiem studenta”, opracowanego przez Digital Care, wynika, że w tym roku akademickim wykorzystanie SI w tym celu planuje co piąty słuchacz wyższej uczelni. Nawet 68 proc. studentów deklaruje,

że będzie używać tego typu narzędzi podczas nauki, głównie jednak do tłumaczeń, tworzenia prezentacji czy pisania prac zaliczeniowych i zadań domowych. 60 proc. badanych ocenia, że SI ma pozytywny wpływ na proces kształcenia. Narzędzia takie jak ChatGPT są popularne także wśród studentów, jednak nie istnieją jeszcze dane na temat skali tego zjawiska. Część studentów obawia się korzystania z narzędzi sztucznej inteligencji do pisania pracy dyplomowej, bo taka praca może zostać uznana przez promotora za niesamodzielną. „Raczej trudno zakładać, że tekst wygenerowany przez sztuczną inteligencję da się rozpoznać ludzkim okiem, dlatego właśnie jest zapotrzebowanie na rozwiązania, takie jak wdrożony moduł detekcji prac generowanych przy użyciu sztucznej inteligencji w Jednolitym Systemie Antyplagiatowym. Natomiast zakładając, że promotor na bieżąco współpracuje ze studentem, zna jego styl pisanie oraz progres, możliwe jest określenie, czy rzeczywiście praca była wykonywana samodzielnie, dlatego ostateczna decyzja zawsze należy tutaj do promotora” – zwraca uwagę Marek Michajłowicz.

Od lutego każda uczelnia może w ramach JSA zbadać, czy dana praca dyplomowa została napisana z użyciem sztucznej inteligencji. „Nowo wdrożony moduł bazuje na hipotezie o tak zwanej regularności tekstu, czyli jego powtarzalności. Żeby dobrze zobrazować schemat działania algorytmu, który został wytrenowany na bardzo dużych zbiorach danych, na bardzo dużych zbiorach tekstu, można się posłużyć miarą o nazwie Perplexity. Dla przykładu tekst wygenerowany w sposób zupełnie losowy cechuje się wysoką wartością Perplexity, natomiast tekst schematyczny, gdzie regularność jest odpowiednio wyższa, ma niski wskaźnik, co skłania nas do uznania za zasadne, że tekst ten mógł zostać wygenerowany za pomocą generatywnych modeli językowych” – wyjaśnia ekspert OPI-PIB.

Jak jednak zaznacza, wspomaganie się sztuczną inteligencją w pisaniu samo w sobie nie jest złe, a uregulowanie tych kwestii jest dopiero postulatem i przedmiotem dyskusji. Narzędzia

takie mogą być nawet korzystne z punktu widzenia możliwości poprawy stylistyki pracy czy wspomagania w doborze literatury przedmiotu, a także produktywności i kreatywności użytkowników. Powinni oni jednak wyraźnie zaznaczać, że dany fragment został napisany przy wsparciu sztucznej inteligencji. „Na bieżąco staramy się uczestniczyć w dialogu ze środowiskiem na temat kierunków rozwoju JSA i dziś rysują dwie główne perspektywy. Pierwszą z nich jest nadążanie za rozwojem sztucznej inteligencji i tekstów generowanych za jej pomocą. Kolejnym istotnym elementem, który od początku wbudowany był w projekt i system, jest powiększanie zbiorów danych, czyli szukanie jak najwięcej baz referencyjnych, które stanowią wiarygodne źródło danych porównawczych na okoliczność plagiatu. Oczywiście cały czas staramy się też wsłuchiwać w głosy środowiska o ergonomii systemu i jego użyteczności” – podkreśla Marek Michajłowicz.

Wdrożenie JSA w 2019 roku i obowiązku weryfikowania w nim wszystkich prac dyplomowych okazało się przełomem na rynku antyplagiatowym i znacząco go uporządkowało. Jeszcze w 2016 roku tylko ok. 30–40 proc. było analizowanych pod kątem plagiatu, a na rynku działało wiele programów opartych na różnych algorytmach i bazach danych. „Od 2019 roku każda praca nadawana przez instytucję przyznającą tytuł zawodowy bądź stopień naukowy musi być zbadana za pomocą tego systemu, który jest dostarczany dobrowolnie. Skorzystało już z niego ponad 400 instytucji, ma on ponad 100 tys. użytkowników” – mówi ekspert OPI-PIB.

Oprócz poszerzania funkcjonalności JSA o kolejne moduły najważniejszym i trwającym nieprzerwanie procesem jest poszerzanie bazy, na której uczą się jego algorytmy. System wykorzystuje zbiory wszystkich pisemnych prac dyplomowych od roku 2009, jak również zasoby zindeksowane polskiego i zagranicznego internetu, który jest tłumaczony w sposób maszynowy. W ciągu czterech lat od uruchomienia JSA za jego pomocą przebadano około 1,5 mln prac dyplomowych. Średnio na

jednego promotora przypada 12 takich prac. Jak podaje OPI, dotychczas ponad 320 tys., które zostały uznane przez system za potencjalne plagiaty, nie zostało zaakceptowanych przez promotorów. W 2022 roku zmniejszyła się skala wykrywanych naruszeń – niespełna 1,7 proc. przebadanych prac (prawie 324 tys.) przekroczyło próg 70 proc. podobieństwa pracy do innego tekstu źródłowego. Średnia dla trzech poprzednich lat to 2,5 proc.

W Polsce konsorcjum naukowo-dydaktyczne Cyber Science, w skład którego wchodzi: Uniwersytet Śląski w Katowicach, Naukowa i Akademicka Sieć Komputerowa – Państwowy Instytut Badawczy NASK, Uniwersytet Ekonomiczny w Katowicach oraz Politechnika Śląska, opracowało rekomendacje dla studentów, doktorantów oraz pracowników badawczych i dydaktycznych w sprawie generatorów treści. Z jednej strony znalazł się w nich apel o włączanie nowych możliwości informatycznych do wachlarza narzędzi wykorzystywanych w procesie dydaktycznym. Z drugiej jednak są tam postulaty mówiące o krytycznym podejściu do źródeł informacji oraz o przeciwdziałaniu wykorzystaniu SI do tworzenia treści, pod którą tylko podpisuje się autor pracy. Jest też rekomendacja wyraźnego wskazania, że dany fragment został wygenerowany przez SI. To tym istotniejsze, że wyniki uzyskane przez generatory treści mogą mijać się z prawdą.

Dyskusja na temat usankcjonowania narzędzi generatywnej sztucznej inteligencji w pisaniu prac naukowych toczy się nie tylko w Polsce. W Holandii, po tym jak władze uczelni z Amsterdamu wykryły, że prace studentów jednej z grup wykładowych zostały napisane za pomocą ChatGPT, doszło do unieważnienia uzyskanych za nie zaliczeń. Plagiat na studiach jest poważnym naruszeniem zasad etycznych i regulaminów akademickich. Jego skutki mogą pociągać za sobą poważne konsekwencje dla studenta, łącznie z wydaleniem ze studiów i niedopuszczeniem do obrony.

Źródło: [Newseria.pl](https://www.newseria.pl)