

# Czatbot

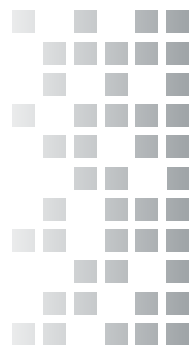
## z siódmej klasy

**Sztuczna inteligencja miała wspomagać edukację i ułatwiać zadanie nauczycielom. Zamiast tego stała się jednym z największych wyzwań dla szkół i uniwersytetów. I na razie nie widać skutecznej metody rozwiązania tego problemu.**

Nie tak miało być. Nauczyciele spodziewali się wsparcia ze strony sztucznej inteligencji w tworzeniu indywidualnych planów nauczania, wyrównywania szans dla dzieci mających trudności z opanowaniem specyficznych partii materiału czy wreszcie uwolnienia od żmudnych zadań administracyjnych. Systemy AI miały wedle tych wyobrażeń sugerować najważniejsze materiały potrzebne do nauki czy chociażby przygotowywać prezentacje dla wykładowców. Mogłyby – uważano jeszcze nie tak dawno – ocenić wypracowanie ucznia i zasugerować poprawki. Co w tym czasie robiłby nauczyciel? Oczywiście miałby więcej czasu na bardziej spersonalizowaną pracę z uczniem, co skutkowałoby wyższym poziomem edukacji i mniejszym stresem dla obu stron.

Inaczej mówiąc: nikt nie miał wątpliwości, że systemy sztucznej inteligencji zrewolucjonizują proces nauczania. Tyle, że ta rewolucja wygląda zupełnie inaczej, niż ją sobie wyobrażaliśmy jeszcze przed rokiem.

Wszystko się zmieniło 30 listopada 2022 r., kiedy OpenAI przedstawiło światu ChatGPT – czatbota wytrenowanego na ogromnych zbiorach danych i zdolnego do udzielania rozbudowanych odpowiedzi na praktycznie dowolne pytania. Po dwóch miesiącach funkcjonowania system sztucznej inteligencji miał już 100 mln użytkowników pytających o wszystko – od polityki i prognozy pogody, po – no właśnie – streszczenia lektur szkolnych i produkcję gotowych wypracowań.



**Piotr Kościelniak**

dziennikarz, popularyzator nauki

Jednocześnie popularność zaczęły zdobywać inne modele AI pozwalające na generowanie treści – m.in. Bard opracowany przez Google czy specjalizujące się w tworzeniu i redagowaniu tekstów Writer, QuillBot, Wordtune, Scholarcy czy Tome, by wymienić tylko kilka najpopularniejszych. W marcu 2023 r. udostępniono również udoskonaloną wersję czatbota OpenAI – ChatGPT Plus opartą na najnowszej generacji modelu GPT-4. Wszystkie te narzędzia pozwalają w ciągu kilku – kilkunastu sekund wygenerować tekst na dowolny temat na poziomie szkolnego wypracowania czy niezbyt zaawansowanego eseju dla studentów uczelni wyższych.



## Szok i rozczarowanie

Tekst generowany przez systemy AI, również w języku polskim, na pierwszy rzut oka niczym nie różni się od tekstu napisanego przez ucznia czy studenta. A to oznacza, że ChatGPT i inne podobne rozwiązania stały się idealnymi narzędziami do oszukiwania. Nie dość, że do uzyskania akceptowalnego wypracowania (czy odpowiedzi testowych) nie jest wymagana żadna wiedza, to w dodatku odpowiedzi bazują na tym, co AI znajdzie w internecie. A zatem do oszustwa dochodzi jeszcze plagiat.

Co ciekawe, jeszcze przed rokiem eksperci specjalizujący się w wykorzystywaniu sztucznej inteligencji w edukacji i szkoleniach zdawali się nie zauważać potencjalnego zagrożenia. W raporcie z 2022 r. „Sztuczna inteligencja (AI) jako megatrend kształtujący edukację. Jak przygotować się na szanse i wyzwania społeczno-gospodarcze związane ze sztuczną inteligencją?” problem ten praktycznie nie występuje<sup>1</sup>. Wybitni eksperci zajęli się w tym dokumencie analizą wpływu AI na proces edukacyjny na różnych poziomach, w tym na poziomie podstawowym i średnim, oczekując, że „takie systemy okażą się bardzo pomocne w automatycznym ocenianiu uczniów, jak również w roli tutorów czyli indywidualnych korepetytorów”. Zaapelowali także o wprowadzenie do programu nauczania podstawowej wiedzy na temat AI i jej praktycznych aspektów.

” *Jak się wydaje, w dziedzinie praktycznego korzystania z AI uczniowie wyprzedzili nauczycieli o kilka długości.*



## Algorytm na trójkę

A możliwości czatbotów potrafią zaskoczyć. Dysponujący dostępem do ogólnej wiedzy (bez precyzyjnego nakierowania na konkretną dziedzinę wiedzy) ChatGPT zdał egzaminy prawnicze na Uniwersytecie Minnesoty oraz biznesowe na Wharton School of Business Uniwersytetu Pensylwanii. Ta sama wersja modelu AI przystąpiła również do egzaminów lekarskich wymaganych do uzyskania licencji medycznej w Stanach Zjednoczonych (USMLE).

Ale uwaga – sztuczna inteligencja zdołała zdać na ocenę dostateczną lub dobrą. Stosując polską skalę ocen, ChatGPT zdał egzaminy prawnicze na 3+, biznesowe na 4, a w przypadku egzaminu lekarskiego zaledwie zbliżył się do wymaganych 60 proc. poprawnych odpowiedzi. Najnowsza wersja ChatGPT radząca sobie z analizą obrazów zaliczyła natomiast egzaminy radiologiczne.

O błyskawicznie rosnącej popularności ChatGPT jako narzędzia ułatwiającego życie studentom niech świadczy statystyka z Uniwersytetu Cardiff.

Podczas styczniowej sesji zaliczeniowej z uniwersyteckiej sieci bezprzewodowej łączono się z ChatGPT 14 443 razy. Miesiąc wcześniej połączeń z systemem OpenAI nie było wcale. Władze uniwersyteckie tłumaczą oficjalnie w BBC, że ten skok to zapewne ruch generowany przez zespoły badaczy chcących przyjrzeć się bliżej nowości – i na pewno nie ma to związku z zaliczeniami semestralnymi.

Jeden ze studentów przyznał jednak, że skorzystał z czatbota do napisania zaliczeniowego wypracowania. Drugie skomponował sam. Za to napisane przez ChatGPT dostał najwyższą możliwą ocenę, natomiast efekt jego samodzielnej pracy oceniono jako zaledwie średni. Inny – że korzystał z systemu sztucznej inteligencji przy większości zadań, jednak nie przepisywał ich słowo w słowo, lecz traktował jako pomoc w usystematyzowaniu wiedzy. Obaj są przekonani, że udział AI jest nie do wykrycia.

Zbliżone wyniki przyniosło badanie w Holandii – kilkuset uczniów przyznało, że korzysta z AI do odrabiania lekcji z języka niderlandzkiego, w tym do poprawiania błędów ortograficznych we własnych pracach oraz pisania wypracowań od podstaw. Również holenderscy uczniowie przekonują, że nauczyciele są nieświadomi takiego procederu i jest on generalnie nie do wykrycia.

Z materiałów dziennikarskich (m.in. „Gazety Wyborczej”) wynika, że sztuczna inteligencja na podobną skalę wykorzystywana jest do oszukiwania nauczycieli również w Polsce. Mało tego, pojawiały się nawet podejrzenia, że istotne raporty tworzone przez naukowców polskich uniwersytetów<sup>2</sup> w rzeczywistości są generowane przez ChatGPT!



## Otwarta wojna...

Co na to sami nauczyciele? Technologia wykrywania oszustw dopiero raczkuje (o czym za chwilę), dlatego szkoły i wykładowcy skazani są na konwencjonalne metody walki z AI. Pierwszy sposób to absolutne wykluczenie możliwości korzystania z AI w jakiegokolwiek formie. Drugi – kreatywne wykorzystanie AI i zmuszenie studentów, aby sami używali takich narzędzi, ale w dokładnie zadany sposób.

<sup>1</sup> <https://kwalifikacje.edu.pl/wp-content/uploads/Sztuczna-inteligencja-jako-megatrend-ksztaltujacy-edukacje.pdf>

<sup>2</sup> <https://wroclaw.wyborcza.pl/wroclaw/7,35771,29726535,skandal-na-uniwersytecie-wroclawskim-raport-za-milion.html>

Zacznijmy od pierwszej metody. Niedługo po udostępnieniu ChatGPT, jeszcze w ubiegłym roku, kilka australijskich uniwersytetów zapowiedziało, że procedury zostaną zmodyfikowane ze względu na możliwość ściągania od maszyny. Zaliczenia będą dokonywane na podstawie egzaminów ustnych lub wypracowań pisanych odręcznie podczas zajęć – bez możliwości skorzystania z komputera.

Tę samą strategię przyjęli wykładowcy prestiżowego University College London. Wcześniej studenci mieli wybór między zaliczeniem w formie praktycznej oceny umiejętności a napisaniem eseju. Teraz zrezygnowano z prac pisemnych. Podobnie postępują szkoły w Stanach Zjednoczonych – wszystkie placówki publiczne w Nowym Jorku, Seattle i Los Angeles zakazały korzystania z ChatGPT.

Dalej poszedł francuski Instytut Nauk Politycznych (Sciences Po), który wprost zakazał studentom posługiwania się systemami w rodzaju ChatGPT do redagowania tekstów. Taki postępek jest traktowany równoznacznie z popełnieniem plagiatu. Władze uczelni zagroziły też karą wydalenia – z wilczym biletem uniemożliwiającym podjęcie jakichkolwiek studiów wyższych we Francji. Z AI można korzystać tylko wówczas, gdy będzie to zainicjowane przez wykładowcę.

### ... czy kohabitacja?

Nie brakuje jednak głosów, że systemy sztucznej inteligencji wykorzystywane w edukacji, zwłaszcza na poziomie akademickim, mogą przyczynić się do lepszej i bardziej twórczej pracy.

– *Wierzę, że w dłuższej perspektywie może to mieć korzystne konsekwencje, doprowadzi do zmiany polityki naukowej i przez to skończy się era generowania nikomu niepotrzebnej intelektualnej papki* – pisze prof. Leszek Pacholski, matematyk, informatyk, były rektor Uniwersytetu Wrocławskiego na łamach „Wszystko co najważniejsze”<sup>3</sup>. Egzaminy i zaliczenia, podczas których AI jest wykorzystywana do oszukiwania, można tak zorganizować, aby uniemożliwić dostęp do internetu i komputerów. Inaczej sprawa wygląda z pracami dyplomowymi – być może konieczne będzie odstąpienie od takiego obowiązku, jeśli promotorzy nie znajdą sposobu na zmuszenie studentów do samodzielnego wysiłku. Obecnie – podkreśla prof. Pacholski – można przecież zamówić napisanie pracy przez wykwalifikowane osoby. ChatGPT sprawi tylko, że te usługi staną się tańsze.

Do kreatywnego wykorzystania systemów AI w nauczaniu zachęca portal WeAre Teachers. – *Dzieci próbujące oszukiwać były w szkołach od zawsze, niezależnie od postępów technologii* – pisze Jill Staake, nauczycielka angielskiego w amerykańskich szkołach średnich. Kluczowym zadaniem

nauczyciela jest jasne postawienie sprawy: *kopiowanie odpowiedzi jest oszustwem, a założenie, że nauczyciel nie umie posługiwać się ChatGPT jest błędne*. Nauczyciele mogą natomiast sami nakłaniać uczniów do korzystania z narzędzi AI, a następnie redagować odpowiedzi i zgłębiać temat przez zadawanie coraz bardziej wymagających pytań.

### Między słowami

Wszystko byłoby łatwiejsze, gdyby nauczyciele mogli wykryć udział maszyny w pracach swoich uczniów. Okazuje się, że obecna generacja czatbotów zostawia „ślady”. Pozwalają one nauczycielowi nabrać podejrzeń, ale nie dadzą nigdy stuprocentowej pewności oszustwa.

Najbardziej oczywisty jest nienaturalny język, pełen powtórzeń i uogólnień (ChatGPT został wytrenowany na materiałach zawierających wiedzę ogólną, lecz brakuje mu dogłębnej znajomości specyficznych tematów). Z tego samego powodu wygenerowany tekst będzie sprawiał wrażenie sklejonego z odrębnych fragmentów – będzie w nim brakować ciągłości myśli i nawiązań do wcześniejszych akapitów.

Zaskakującym sposobem na odkrycie udziału AI jest poszukiwanie błędów interpunkcyjnych, gramatycznych i literówek. Takie błędy zdarzają się każdemu, a ich brak oznacza niemal na pewno, że tekst został wygenerowany przez maszynę i przeklejony (z przepisany) przez ucznia.

A co z rozwiązaniami bazującymi na technologii? Nauczyciele mogą oczywiście skorzystać z kilku narzędzi do wykrywania plagiatów (Copyscape, Grammarly, Copywritely czy Plagiat.pl), a promotorzy mają dostęp do Jednolitego Systemu Antyplagiatowego, jednak programy te nie zostały opracowane z myślą o wykrywaniu udziału sztucznej inteligencji. Takie narzędzie opracowała jednak sama OpenAI.

AI Text Classifier jest specjalnym modelem GPT, którego zadaniem jest stwierdzenie czy tekst napisała sztuczna inteligencja (dowolna, nie tylko ChatGPT). Ma on liczne ograniczenia, spośród których najpoważniejsze jest takie, że został wytrenowany z języku angielskim i z innymi językami radzi sobie słabiej.

Fragment tekstu, który Państwo właśnie czytacie, został oceniony jako „pochodzenie niepewne”.

W internecie dostępne są również inne narzędzia, m.in. ZeroGPT, chatgptdetector czy AI Content Detector, które

<sup>3</sup> <https://wszystkoconajwazniejsze.pl/prof-leszek-pacholski-chatgpt-nie-zrewolucjonizuje-nauki/>

mają wykrywać tekst wygenerowany przez AI. Mój testowy fragment niniejszego tekstu został oceniony odpowiednio: jako całkowicie ludzki, w 31 proc. wygenerowany przez AI i wreszcie jako niemożliwy do zbadania. Dlatego efektywność podobnych rozwiązań w wykrywaniu oszustw trzeba dziś określić jako niską.

Jednym z rozważanych obecnie przez OpenAI pomysłów jest wprowadzenie swego rodzaju znaku wodnego pozwalającego zidentyfikować tekst jako pochodzący z generatora ChatGPT. Wedle zapowiedzi miałyby to być specyficzne pseudolosowe ciągi słów niezaburzający sensowności tekstu i niemożliwy do wychwycenia na pierwszy rzut oka.

– *Chcemy, aby treść generowana przez systemy AI nie mogła uchodzić za treści tworzone przez człowieka. To będzie pomocne między innymi w walce z plagiatami prac akademickich, ale również zapobiegnie szerzeniu masowo generowanej propagandy politycznej* – mówi prof. Scott Aaronson, jeden z naukowców zespołu OpenAI. Według Aaronsona, firma ma już działający prototyp algorytmu dopisującego znak wodny. Ma on działać jako kolejna warstwa na gotowym tekście. Słowa i znaki mają zostać odpowiednio zmodyfikowane w taki sposób, aby nie zmieniając sensu oznaczyć tekst jako wygenerowany sztucznie. Żeby wykryć ten znak wodny, niezbędny byłby jednak specjalny klucz pozwalający wyszukiwać „ślady AI” w badanym materiale.

Wcześniej podobny pomysł zaprezentowali niemieccy naukowcy z CISPA Helmholtz Center for Information Security. Zaproponowali oni technikę podstawiania synonimów i stosowania specyficznych zasad składniowych, które pozwoliłyby na identyfikację treści pisanych przez AI.

Problem z takimi rozwiązaniami polega jednak na tym, że tekst wygenerowany przez jeden system AI można następnie przepuścić przez inny, niezostawiający znaków wodnych, z żądaniem przeredagowania. Niezbyt długie materiały – wypracowania szkolne czy nawet krótsze eseje na poziomie akademickim – można również przeredagować ręcznie, unikając tym samym wykrycia. Jak argumentują inni naukowcy (m.in. Jack Hessel z Allen Institute for AI), zbyt intensywne użycie metod znakowania doprowadziłoby do oczywistych – widocznych – błędów językowych, a zbyt subtelne zmiany mogłyby zostać niewykryte lub nie dawać stuprocentowej pew-

ności fałszerstwa. Technika znakowania kryptograficznego dla ChatGPT proponowana przez OpenAI oznaczałaby zaś, że jedyną organizacją mogącą ustalić, czy tekst pochodzi z generatora byłoby... samo OpenAI. Dla nauczycieli sprawdzających prace domowe taka technika byłaby zaś nieosiągalna.



### Problem nie tylko w szkołach

Wyzwaniem, jakim jest kopiowanie gotowych odpowiedzi i rozwiązań z ChatGPT, zajęły się również instytucje nieakademickie. Serwis dla programistów Stack Overflow zakazał używania narzędzi AI do generowania odpowiedzi na pytania społeczności. Tu nie chodziło jednak o sam fakt plagiatowania innych, ale o to, że odpowiedzi udzielane przez sztuczną inteligencję często były nieprawdziwe. Przybywa również firm, które zakazują pracownikom używania ChatGPT – zasady takie wprowadziły m.in. japońskie Softbank, Fujitsu i Panasonic, powołując się na konieczność ograniczenia ryzyka wycieku danych.

Zagrożenia dla celowości edukacji to tylko jedna z kontrowersji otaczających systemy generatywnej sztucznej inteligencji. Problemy dotyczą m.in. ochrony praw autorskich i naruszania prywatności, a także potencjalnych zagrożeń fake newsami, manipulacją oraz zastępowania na rynku pracy ludzi maszynami. To powody, dla których rozwiązania te zostały tymczasowo zablokowane we Włoszech (z ChatGPT można ponownie korzystać tam od 29 kwietnia) i częściowo w Irlandii. Unia Europejska przygotowuje również nowe regulacje prawne dotyczące transparentności systemów AI, odpowiedzialności jej twórców, wykorzystania chronionych prawem utworów do szkolenia algorytmów itp. Zachodnie systemy sztucznej inteligencji są oczywiście zakazane w Rosji, Chinach, Iranie i Korei Płn.

Zagrożenia ze strony niekontrolowanego rozwoju, a przede wszystkim nieodpowiedzialnego stosowania AI dostrzegają sygnatariusze listu otwartego Future of Life Institute. Domagają się oni natychmiastowego wprowadzenia sześciomiesięcznego moratorium na eksperymenty z kolejnymi generacjami systemów sztucznej inteligencji, przestrzegając przed katastrofalnymi efektami społecznymi wdrażania coraz doskonalszych czatbotów<sup>4</sup>. Pod listem podpisali się m.in. współzałożyciel Apple Steve Wozniak i Elon Musk.

<sup>4</sup> <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>