

# Wskaż poprawną odpowiedź: Prawda czy AI

AGNIESZKA BOROWIECKA • BARTŁOMIEJ KROWIAK

Nauczyciele konsultanci

Ośrodek Edukacji Informatycznej i Zastosowań Komputerów w Warszawie

*Czasami tylko nam się wydaje, że coś jest prawdziwe,  
choć wcale takie nie jest.*

Moony Witcher<sup>1</sup>, *Nina i Tajemnica Ósmej Nuty*

Współczesny internet jest niczym labirynt. Media społecznościowe oraz portale informacyjne zasympowane są nowymi treściami, wśród których coraz więcej zostało wygenerowanych przy użyciu zaawansowanych narzędzi sztucznej inteligencji. Grafiki, zdjęcia i krótkie filmy wideo są obecnie tworzone z taką precyzją, że często trudno je odróżnić od autentycznych materiałów.

Należy podkreślić, że problem dezinformacji nie leży w samej technologii. **Sztuczna inteligencja to jedynie narzędzie**, podczas gdy za każdym podrobionym obrazem lub sfalszowanym filmem stoi człowiek używający AI w konkretnym celu – często do manipulacji, oszustwa lub celowego wprowadzania w błąd. Konsekwencje mogą być poważne: od dezinformacji politycznej, przez oszustwa finansowe, aż po wywoływanie paniki społecznej. To właśnie te ludzkie intencje stoją za zjawiskiem tak zwanych deepfake'ów, czyli fałszerstw audiowizualnych łudząco przypominających prawdziwe nagrania.

Korzystanie z narzędzi sztucznej inteligencji bez złych intencji również nie jest w pełni bezpieczne, możemy się bowiem spotkać ze zjawiskiem zwanym „halucynowaniem”. Występuje ono wtedy, gdy

model językowy podaje błędne cytaty i fakty albo tworzy nieistniejące informacje. Modele AI działają w oparciu o zasady statystyki i teorii prawdopodobieństwa – ich celem jest stworzenie najbardziej prawdopodobnej sekwencji słów lub pikseli na podstawie ogromnych zbiorów danych, a nie odzwierciedlenie faktycznej rzeczywistości. Jest to logiczne z punktu widzenia programu, lecz może okazać się zgubne dla użytkownika poszukującego rzetelnej informacji.

Grafika i wideo to najbardziej angażujące formy dezinformacji. Istnieją jednak praktyczne i proste metody, które pozwalają sprawnie weryfikować treści i odpowiedzieć na pytanie: Prawda czy AI?

## ZWERYFIKUJ TO SAMODZIELNIE!

Internet coraz bardziej zalewa fala treści generowanych przy pomocy sztucznej inteligencji, dlatego zapoznając się z interesującymi nas informacjami, nie można polegać wyłącznie na pierwszym wrażeniu. Najważniejszym narzędziem weryfikacji

<sup>1</sup> Moony Witcher – pseudonim artystyczny włoskiej pisarki, dziennikarki i scenarzystki Robertny Rizzo, autorki popularnych cykli literatury fantastycznej dla młodzieży.

## WSKAŹ POPRAWNĄ ODPOWIEDŹ: PRAWDA CZY AI

grafiki, wideo czy tekstu nie jest żaden program komputerowy, ale krytyczne myślenie i chłodny osąd. Należy zdystansować się od treści i ocenić ją bez emocji, ponieważ twórcy często celowo wykorzystują chwytliwe tematy i sensacyjne obrazy, aby wywołać szybką, nieprzemyślaną reakcję. Spokój i krytyczne myślenie to nasza pierwsza linia obrony.

W świecie pełnym złożonych mechanizmów dezinformacji niezbędne jest opanowanie podstawowych metod weryfikacji treści. Istnieje mnóstwo narzędzi oferujących płatne lub bezpłatne funkcje sprawdzające, które czasem wymagają dodatkowo założenia konta. Nauczycielom i uczniom potrzebne są jednak rozwiązania dostępne od ręki. Przyjrzyjmy się dwóm najłatwiejszym i najbardziej dostępnym sposobom: analizie wizualnej oraz wsteczemu wyszukiwaniu obrazem.

Jeżeli zdjęcie wydaje się podejrzane, należy się mu uważnie przyjrzeć. Nawet najbardziej zaawansowane generatory grafiki wciąż popełniają błędy, tzw. artefakty AI, które pozwalają rozpoznać pochodzenie obrazu. Model nie kopiuje rzeczywistości, lecz rekonstruuje obraz na podstawie danych statystycznych. Często oznacza to, że AI świetnie radzi sobie z centralnym, głównym obiektem, ale detale tła czy mniej istotne szczegóły obrazu generuje z mniejszą dokładnością lub logiką.

Typową bolączką generatorów pozostają ręce: pojawia się nadmiarowa liczba palców, zbyt długie kciuki, zniekształcone stawy czy dłonie nienaturalnie zlewające się z ramieniem lub trzymanym przedmiotem<sup>2</sup>. Podobnie, nawet jeśli główne elementy twarzy są realistyczne, często można zauważyć asymetrię oczu, zwłaszcza u osób w tle. Również zęby bywają złane w jednolitą, nienaturalnie białą masę.



**Ilustracja 1.** Grafika wygenerowana przy użyciu AI – krawat mężczyzny jest naturalnie położony, ale pozycja dłoni wokół teczki wydaje się podejrzana<sup>3</sup>

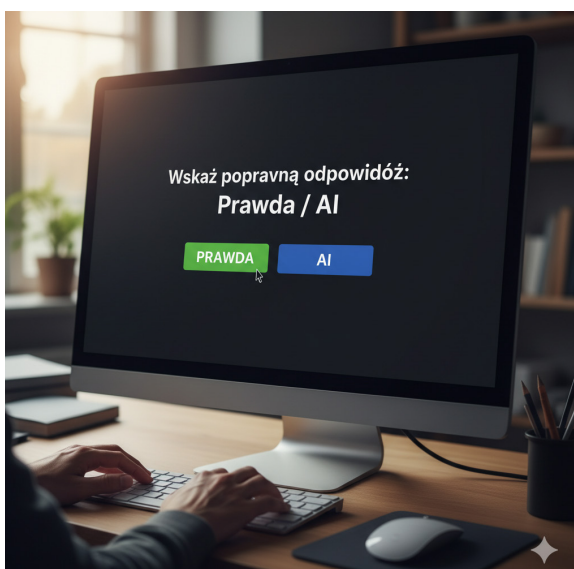
<sup>2</sup> K. Wesolowski, *Jak poznać zdjęcia generowane przez sztuczną inteligencję?*, <https://tinyurl.com/ytzxfy9m>

<sup>3</sup> *5 Telltale Signs That a Photo Is AI-generated*, <https://tinyurl.com/56svhbas>

## AGNIESZKA BOROWIECKA BATŁOMIEJ KROWIAK

Kolejny obszar weryfikacji to tło i logika sceny. Analiza tła często ujawnia błędy w perspektywie, niespójność cieni, a obiekty mogą się powtarzać w dziwny, zapętłony sposób. AI ma również problemy z tekstem – dużo napisów, logotypów czy etykiet w tle jest z reguły zniekształconych i składających się z losowych lub nieczytelnych liter.

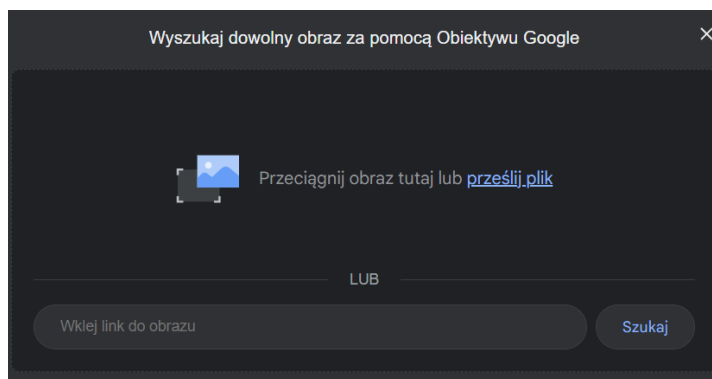
Ostatnim sygnałem ostrzegawczym bywa tekstura – nienaturalna gładkość skóry lub powtarzalność wzorów mogą wskazywać na interwencję AI.



**Ilustracja 2.** Grafika stworzona przy pomocy generatora AI, która miała zawierać napis z tytułem artykułu

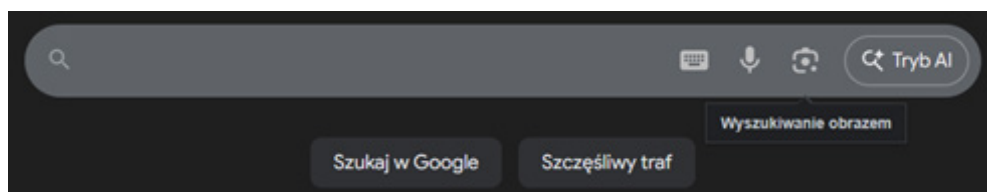
Jeżeli wstępna analiza wizualna nie rozwieje wątpliwości, kolejnym krokiem może być sprawdzenie historii obrazu. Celem wyszukiwania wstecznego jest ustalenie, kiedy i gdzie dane zdjęcie lub grafika pojawiło się w sieci po raz pierwszy, a także w jakim kontekście było i jest używane. Najprostszym i najczęściej wykorzystywanym narzędziem jest *Wyszukiwanie obrazem* w wyszukiwarce Google lub Google Grafika<sup>4</sup>.

Proces weryfikacji jest prosty – aby rozpocząć, wystarczy zapisać podejrzany o wygenerowanie przez AI plik graficzny na swoim komputerze lub telefonie, a następnie otworzyć wyszukiwarkę Google, kliknąć ikonę aparatu (Ilustracja 3) i wgrać zapisane zdjęcie z dysku. Można również wkleić link do obrazu dostępnego w sieci, przeciągnąć obraz z innego okna lub wkleić grafikę ze schowka.



**Ilustracja 4.** Panel służący do wgrzywania zdjęć w *Wyszukiwaniu obrazem* Google

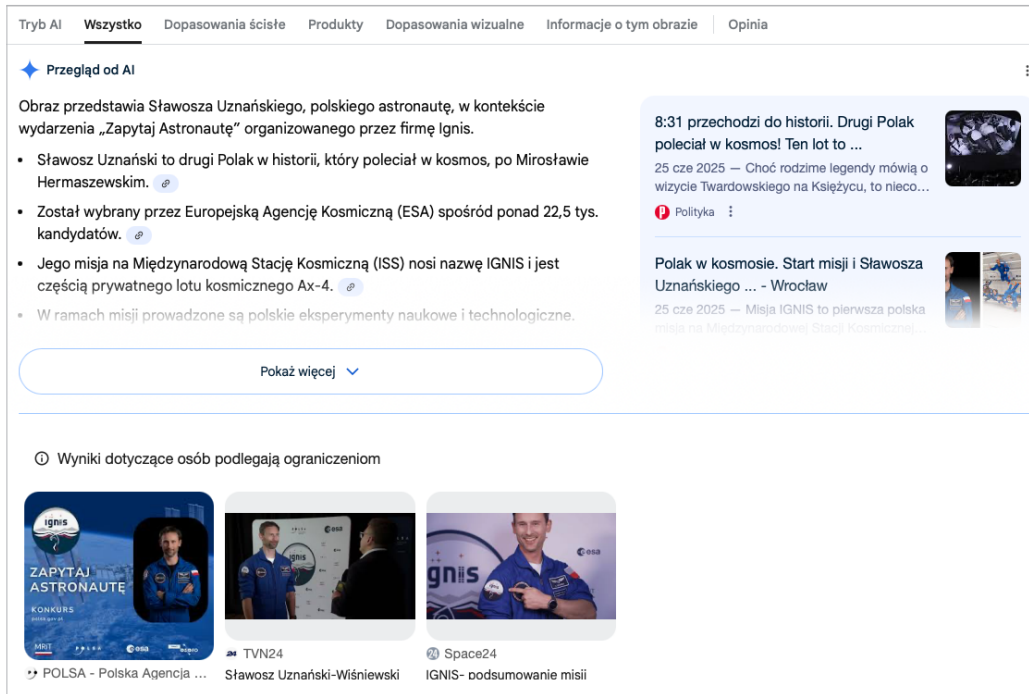
Google wyświetli wszystkie podobne lub identyczne obrazy, w wielu sytuacjach zobaczymy również wygenerowany przez AI opis zawartości wczytanej grafiki. Zakładka *Dopasowania ściste* pozwala obejrzeć identyczne zdjęcia wraz z linkami do zawierających je stron, *Dopasowanie wizualne* – podobne zdjęcia znalezione w internecie, a *Informacje o tym obrazie* podają listę stron z podobną grafiką oraz informacje o tym, jak długo jest ona dostępna w sieci. W przypadku oryginalnych zdjęć możemy zobaczyć komunikat „Nie znaleziono dopasowań ścistych”. Może to oznaczać, że obraz jest unikalny lub nie został jeszcze szeroko udostępniony.



**Ilustracja 3.** Pasek wyszukiwania przeglądarki Google Chrome wraz z ikoną dotyczącą szukania obrazem

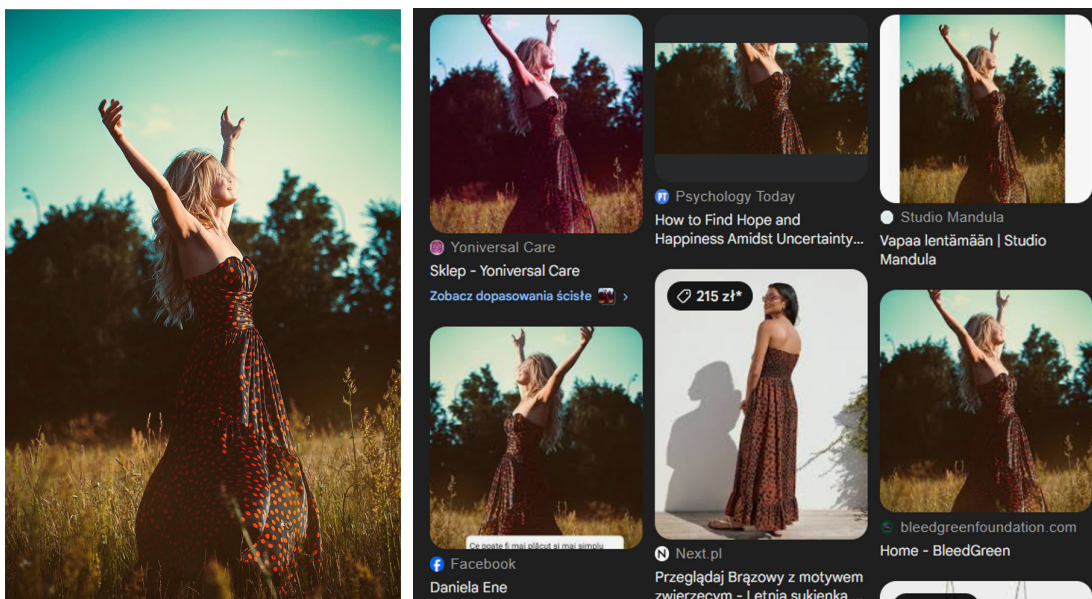
<sup>4</sup> <https://images.google.com>

## WSKAŹ POPRAWNĄ ODPOWIEDŹ: PRAWDA CZY AI



Ilustracja 5. Wyniki wyszukiwania obrazem Google

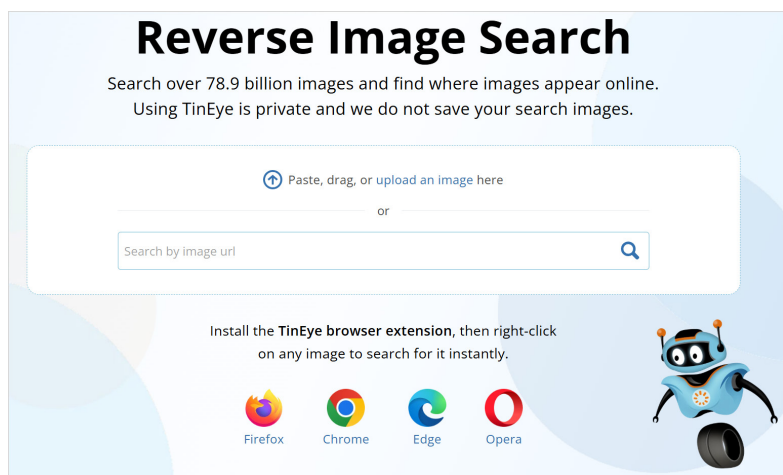
Często okazuje się, że obraz powiązany jest z zupełnie inną historią, został wykorzystany na dziesiątkach kont fałszywych profili lub jest ilustracją stockową<sup>5</sup> o innym tytule. Wyszukiwanie obrazem pozwala na demaskowanie oszustw na kontach w mediach społecznościowych. Wgranie awatara w wyszukiwarce może ujawnić, że na zdjęciu widnieje osoba publiczna z innego kraju albo ta sama twarz została wykorzystana przez wiele różnych kont „inwestorów” lub innych internetowych naciągaczy.



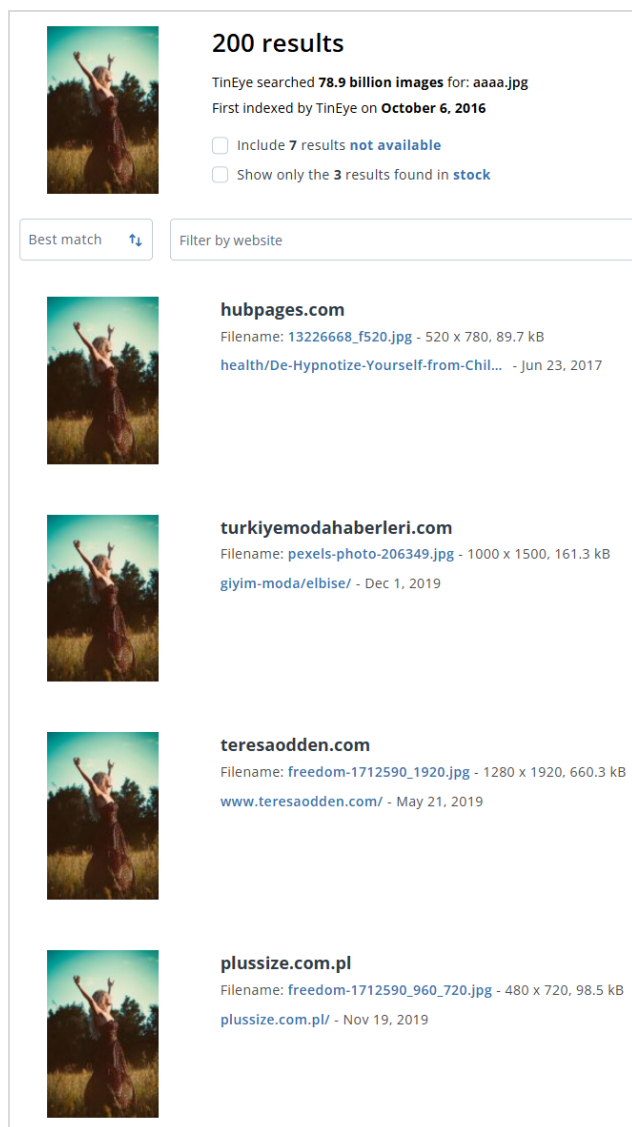
Ilustracja 6. Grafika stockowa<sup>6</sup> oraz wynik wyszukiwania obrazem Google

<sup>5</sup> Ilustracja stockowa – gotowy obraz lub grafika dostępna do wykorzystania na podstawie określonej licencji w bankach zdjęć, takich jak Shutterstock lub Adobe Stock.

<sup>6</sup> <https://tiny.pl/d74bqsvm>

AGNIESZKA BOROWIECKA  
BATŁOMIEJ KROWIAK

Ilustracja 7. Widok na stronie głównej TinEye wraz z polem do wprowadzenia grafiki



Ilustracja 8. Efekt pracy TinEye na podstawie poprzedniego zdjęcia

Choć Google ma olbrzymią bazę danych, czasami algorytmy nie są w stanie odnaleźć pierwotnego źródła obrazu. Potrzebne jest wtedy wsparcie dodatkowe niezależnego narzędzia, takiego jak **TinEye**<sup>7</sup> – wyspecjalizowana wyszukiwarka, która przeszukuje własną bazę obrazów.

Proces weryfikacji w TinEye jest analogiczny, wystarczy wgrać plik lub wkleić adres obrazu. Narzędzie oferuje jednak inną cenną funkcjonalność: bardzo precyzyjne filtrowanie wyników według daty publikacji. Można

posortować wyniki, aby zobaczyć najstarsze użycie danego obrazu w sieci, a także sprawdzić, jak grafika zmieniała się w czasie. Jeżeli zdjęcie przedstawiające hipotetycznie nowe wydarzenie było już w obiegu kilka lat temu albo zostało celowo wykadrowane/odwrócone, by pasowało do faksztywnej narracji, TinEye z łatwością to wychwyci.

Wyszukiwarka TinEye w połączeniu z Obiektywem Google i krytycznym myśleniem stanowi solidny i niezawodny fundament dla każdego, kto chce skutecznie weryfikować cyfrowe treści.

## AI ZDEMASKOWANE PRZEZ... AI?

Proste metody weryfikacji dostępne dla każdego stają się niekiedy niewystarczające. Generatory sztucznej inteligencji rozwijają się w zawrotnym tempie. To, co do niedawna było łatwym do wykrycia artefaktem AI w postaci na przykład nienaturalnie wygiętej dłoni, obecnie jest już często korygowane niemal perfekcyjnie. W efekcie zdolność ludzkiego oka do wykrycia fałszerstwa drastycznie maleje.

Na szczęście rozwój generatywnej AI napędza również segment oprogramowania mającego demaskować nadużycia związane z niewłaściwym jej wykorzystaniem. Powstaje coraz więcej specjalistycznych skanerów, które używają własnych algorytmów sztucznej inteligencji, aby wykryć cyfrowe ślady pozostawione przez inne narzędzia. Jest to

<sup>7</sup> <https://tineye.com>

## WSKAŹ POPRAWNĄ ODPOWIEŹ: PRAWDA CZY AI

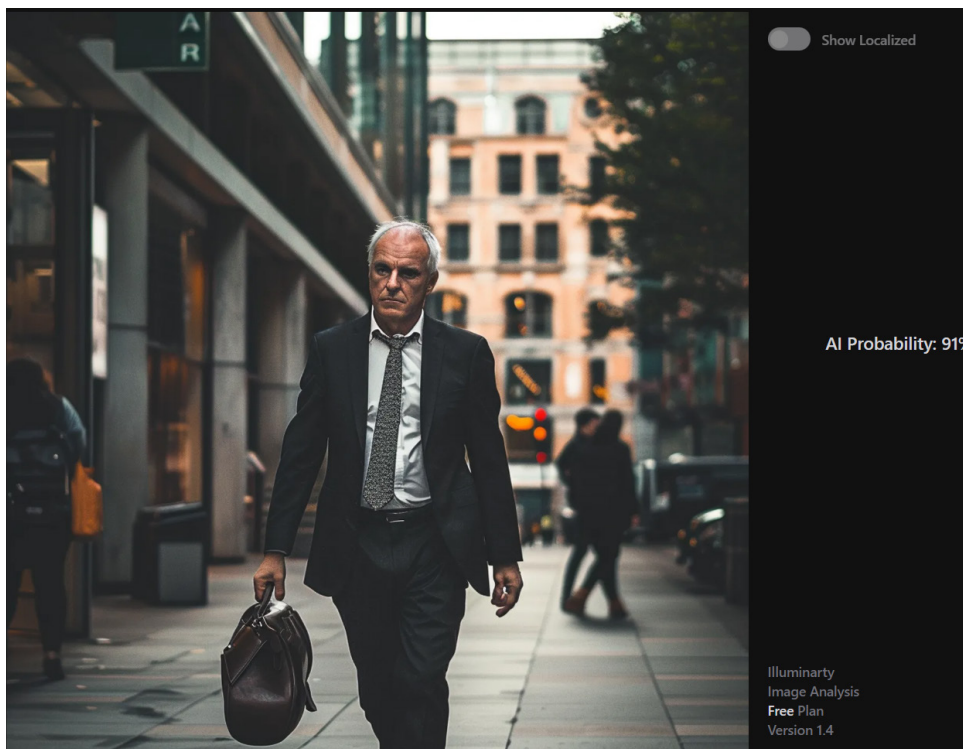
swoista walka technologii z technologią, jednak warto pamiętać, że AI generujące treści ma przewagę w tempie rozwoju, a technologia skanująca nigdy nie będzie oferować stuprocentowej skuteczności. Mimo to takie narzędzia stanowią profesjonalne wsparcie w trudnych przypadkach.

Do szybkiego i szczegółowego sprawdzania obrazów mogą posłużyć specjalne narzędzia online, takie jak:

- **illuminarty.ai**<sup>8</sup> – popularna strona internetowa do weryfikacji obrazów oferująca zaawansowaną analizę techniczną. Umożliwia wykrywanie, czy obraz został wygenerowany przez AI lub był w pewnym stopniu modyfikowany. Co istotne – platforma ta jest bezpłatna w podstawowej wersji i dostępna bez konieczności zakładania konta, co czyni ją łatwo dostępną dla każdego.
- **AI or Not**<sup>9</sup> – skaner oferujący szybką weryfikację obrazu i oceniający prawdopodobieństwo, czy grafika jest dziełem ludzkim czy maszynowym. Decydując się na jego użycie, należy pamiętać o konieczności zalogowania się lub założenia konta, co w niektórych sytuacjach może stanowić pewną barierę. Wersja bezpłatna pozwala ocenić do 10 obrazów lub przeanalizować tekst o długości do 5000 słów.

Materiał wideo stanowi znacznie większe wyzwanie dla weryfikacji. Tworzenie i rozpowszechnianie deepfake'ów jest zatem jednym z najpoważniejszych zagrożeń dezinformacyjnych.

Przykładowym narzędziem do analizy tego typu materiałów jest **Deepware Scanner**<sup>10</sup>. To platforma pozwalająca na wczytanie linku do pliku wideo w celu jego przeskanowania i określenia, czy nosi ślady manipulacji AI. Tak jak w przypadku illuminarty, korzystanie z Deepware nie wymaga zakładania konta, dzięki czemu jest to wygodne narzędzie służące szybkiej weryfikacji filmów.

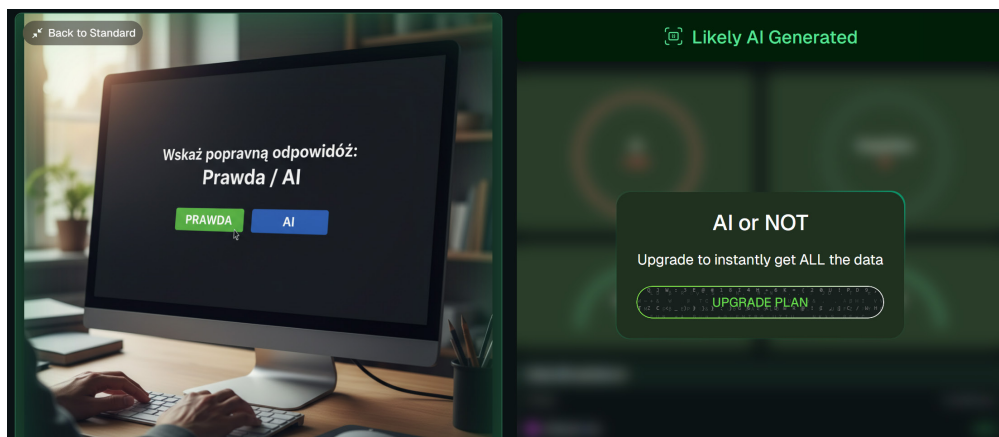


Ilustracja 9. Efekt pracy illuminarty.ai na podstawie zdjęcia z początku artykułu

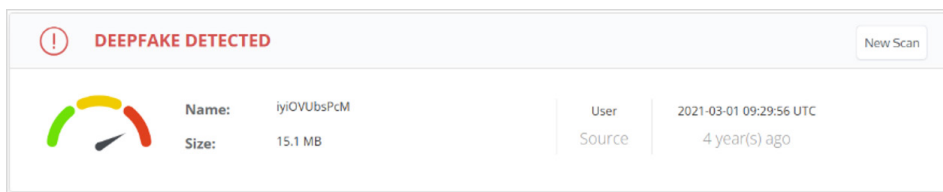
<sup>8</sup> <https://app.illuminarty.ai>

<sup>9</sup> <https://www.aiornot.com>

<sup>10</sup> <https://scanner.deepware.ai>

AGNIESZKA BOROWIECKA  
BATŁOMIEJ KROWIAK

Ilustracja 10. Efekt pracy AlorNot.com na podstawie wgranego zdjęcia



Ilustracja 11. Przykład działania narzędzia Scanner Deepware

Choć przedstawione narzędzia wykorzystują zaawansowane algorytmy do analizowania materiałów audiowizualnych, nie można traktować ich werdyktu jako ostatecznej i niepodważalnej prawdy. Zawsze powinniśmy postępować z dużą ostrożnością. Gdy skaner wskaże, że obraz jest czysty, ale chłodny osąd nadal sugeruje fałszerstwo, nie należy pozbywać się wątpliwości. W takich sytuacjach wkracza trzecia linia obrony: wsparcie ekspertów.

## ZATRUDNIJ FACHOWCA

Weryfikacja treści cyfrowych, choć możliwa na własną rękę czy przy wsparciu skanerów AI, czasem okazuje się zbyt trudna. Nawet najbardziej zaangażowany czytelnik może nie mieć czasu na czasochłonną analizę źródła, która dodatkowo może wymagać dostępu do specjalistycznych baz danych i narzędzi niedostępnych dla przeciętnego internauty. Właśnie w takich sytuacjach konieczne staje się zwrócenie do profesjonalistów.

Możemy szukać wsparcia zorganizowanych grup, które specjalizują się w metodycznej i bezstronnej weryfikacji informacji, tak zwanych grup

fact-checkingowych<sup>11</sup>. Są to zróżnicowane podmioty, od stowarzyszeń non-profit, przez dziennikarskie zespoły redakcyjne, po instytucje naukowe. Ich pierwotnym celem było sprawdzanie prawdziwości wypowiedzi polityków oraz weryfikacja wiadomości z mediów głównego nurtu. Obecnie ich działalność jest dodatkowo związana z walką z dezinformacją generowaną przy użyciu narzędzi sztucznej inteligencji. Coraz więcej wytworów AI przedostaje się do masowych mediów, stąd eksperci musieli rozszerzyć swoje działania o zaawansowaną analizę obrazu i wideo. Ich rola polega nie tylko na wskazaniu, co jest fałszem, ale także udokumentowaniu procesu weryfikacji, co ma ogromną wartość edukacyjną dla społeczeństwa.

W Polsce działa kilka kluczowych podmiotów, które zajmują się tego typu zagadnieniami, jak na przykład Stowarzyszenie Demagog<sup>12</sup>, Konkret24<sup>13</sup>, AFP Sprawdzam<sup>14</sup> czy instytut badawczo-rozwojowy NASK<sup>15</sup>.

<sup>11</sup> *Czym jest fact-checking? – zarys inicjatyw na świecie i w Polsce*, <https://tinyurl.com/3da8h3fa>

<sup>12</sup> <https://demagog.org.pl>

<sup>13</sup> <https://konkret24.tvn24.pl>

<sup>14</sup> <https://sprawdzam.afp.com>

<sup>15</sup> <https://www.nask.pl>

## WSKAŹ POPRAWNĄ ODPOWIEŹ: PRAWDA CZY AI

Problem fałszywych treści ma charakter globalny. Istnieje wiele zagranicznych grup fact-checkingowych, które współpracują w ramach międzynarodowych sieci, co zapewnia spójność standardów i wymianę wiedzy w zakresie wykrywania najbardziej zaawansowanych fałszerstw AI.

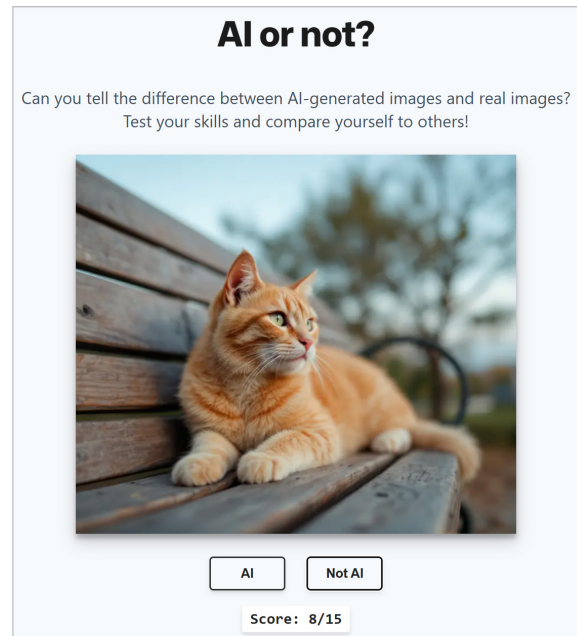
### NAUKA POPRZEZ ZABAWĘ

Umiejętność weryfikowania treści w internecie to kluczowa kompetencja XXI wieku, a najlepszym sposobem na jej rozwijanie, zwłaszcza w pracy z uczniami, jest łączenie nauki z zabawą. Gry i proste quizy przeglądarkowe to efektywny mechanizm gamifikacji, który pozwala wyczerpać percepcję, wzmocnić krytyczne myślenie i wykształcić nawyk szybkiego rozpoznawania cyfrowych oszustw.

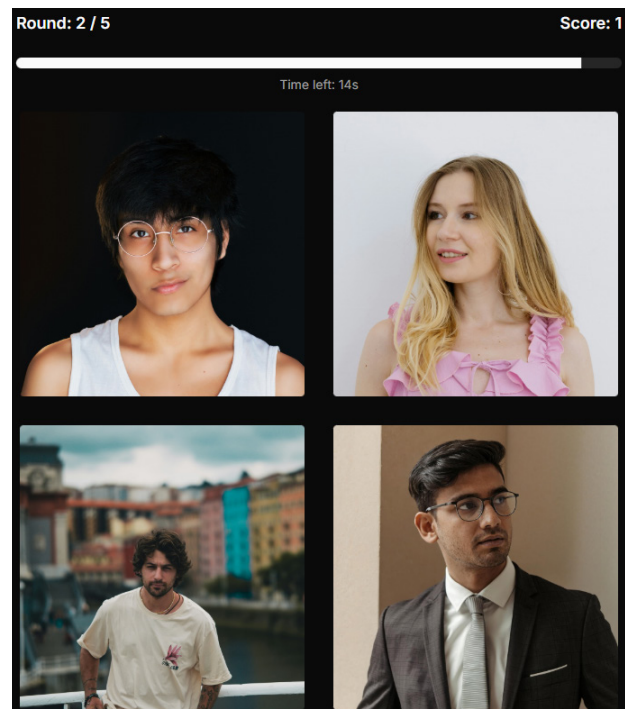
Jedną z takich witryn jest **Sight Engine: AI or Not**<sup>16</sup>. Dostępne są tu testy pozwalające zidentyfikować obrazy lub muzykę wygenerowaną przez AI. W teście wizualnym graczowi prezentowany jest obraz, a zadaniem jest zdecydowanie, czy jest on prawdziwym zdjęciem, czy też został wygenerowany przez sztuczną inteligencję. Gra wymaga skupienia się na detalach – analizie dłoni, zębów, zniekształceń tła. Mimo natychmiastowego werdyktu po zakończeniu rundy gracz nie otrzymuje żadnej rozbudowanej informacji zwrotnej, która uzasadniałaby uzyskany wynik lub opisywała popełnione błędy. W teście muzycznym odtwarzane są kilkudziesięciosekundowe pliki audio, na podstawie których gracz powinien rozstrzygnąć, czy ich autorem jest człowiek, czy sztuczna inteligencja. Ponownie nie otrzymujemy żadnej dodatkowej informacji poza stwierdzeniem, czy nasza odpowiedź była poprawna, czy nie.

W grze **Real or Fake**<sup>17</sup> gracz w wyznaczonym przez siebie czasie ma za zadanie wskazać jedną wygenerowaną przez AI grafikę spośród czterech prezentowanych. Po wskazaniu wybranego rysunku system informuje, czy odpowiedź była prawidłowa. W przypadku błędnej odpowiedzi dodatkowo

wskazywany jest obraz wygenerowany przez AI. To doświadczenie jest cenną lekcją – uświadamia, jak łatwo możemy ulec złudzeniu i jak wiele rzekomo prawdziwych zdjęć w mediach faktycznie może być tworem AI.



Ilustracja 12. Jedno z zadań w grze przeglądarkowej Sight Engine: AI or Not



Ilustracja 13. Jedno z zadań w quizie Real or Fake

<sup>16</sup> <https://sightengine.com/ai-or-not>

<sup>17</sup> <https://tiny.pl/spsxb9cx>

## AGNIESZKA BOROWIECKA BATŁOMIEJ KROWIAK

Inny charakter ma **Detect Fakes**<sup>18</sup>, projekt naukowy stworzony przez ekspertów z Kellogg School of Management na Northwestern University. Celem autorów projektu jest zrozumienie, jak dobrze ludzie potrafią odróżnić obrazy prawdziwe od stworzonych przez sztuczną inteligencję. Prezentowana jest seria zdjęć, a naszym zadaniem jest ocena, czy prezentowane zdjęcie jest prawdziwe, czy zostało stworzone przy użyciu AI, a następnie określenie, w jakim stopniu jesteśmy pewni swojej odpowiedzi. Mamy także możliwość uzasadnienia swojego wyboru. Najważniejszy element następuje przed ostatecznym werdyktem – prezentowane są wskazówki, które zachęcają do ponownej analizy kontekstu i detali. Dopiero po potwierdzeniu lub zmianie zdania, otrzymujemy pełną informację o faktycznym statusie zdjęcia. Takie podejście do weryfikacji zdjęć stanowi cenną lekcję dla nauczycieli i młodzieży.

### PODSUMOWANIE

Problem niewłaściwego wykorzystywania treści generowanych przez sztuczną inteligencję stanowi

istotne wyzwanie cywilizacyjne. Skuteczna obrona przed fałszem audiowizualnym opiera się na kilku poziomach weryfikacji. Proces ten rozpoczyna się od samodzielnej krytycznej analizy – poszukiwania artefaktów AI, stosowania wyszukiwania wstecznego (np. Google i TinEye) oraz zachowania chłodnego osądu.

Gdy to podejście jest niewystarczające, należy sięgnąć po specjalistyczne narzędzia wykorzystujące własne algorytmy, które pomagają w wykryciu ukrytych śladów cyfrowej manipulacji. W przypadku najbardziej zaawansowanych i złożonych fałszerstw ostatecznym etapem jest zwrócenie się do profesjonalistów, którzy dysponują fachową wiedzą i dostępem do niezbędnych baz danych.

Sztuczna inteligencja pozostaje potężnym narzędziem, ale to człowiek decyduje o celu jej użycia. Kluczem do przetrwania w labiryncie cyfrowych treści jest nieustająca edukacja, utrzymywanie dystansu do sensacyjnych doniesień i promowanie świadomego, krytycznego podejścia do każdej informacji. W środowisku pedagogicznym zdobyta wiedza powinna służyć przede wszystkim kształtowaniu cyfrowej świadomości. Warto stosować metody aktywizujące, takie jak gry przeglądarkowe, które doskonałą czujność. Niezbędne jest, aby podczas pracy z uczniami skupiać się na nauczaniu metody weryfikacji, a nie tylko na samych narzędziach. ●


### AI-Generated or Real?

Please identify whether the image is real or fake and how confident you are. Scroll Down.

↓

**Advice:**

- Look at the details of the suit and tie—do the patterns and folds appear natural?
- Check the fingers on the hand pointing—do they look properly formed and positioned?
- Pay attention to the way the saxophone is held—does the grip and posture match how the instrument is typically played?



### Real or AI-Generated?

**Real:** This is a real image.

**Fake:** This is an AI-generated image.

**How confident are you?**

Not at all

Slightly

Moderately

Very

Perfectly

**How much do you think others would agree with your judgment?**

Almost no one

Few

About half

Most

Almost everyone

Submit your final guess

Ilustracja 14. Wygląd jednego z zadań wraz z późniejszymi podpowiedziami na stronie Detect Fakes

<sup>18</sup> <https://detectfakes.kellogg.northwestern.edu>