„Wyostrzone” nie znaczy prawdziwe

O tym jak narzędzia AI mogą zmieniać szczegóły zdjęć i jak na własne życzenie wprowadzamy szum informacyjny.



NASK

„Wyostrzone” nie znaczy prawdziwe

O tym jak narzędzia AI mogą zmieniać szczegóły zdjęć i jak na własne życzenie wprowadzamy szum informacyjny.

Adrian Kordas

Ewelina Bartuzi-Trokielewicz

Katarzyna Roszczewska

NASK

Ośrodek Analizy Dezinformacji NASK

PROJEKT FINANSOWANY ZE ŚRODKÓW MINISTERSTWA CYFRYZACJI

CWZ CENTRUM WALKI Z DEZINFORMACJĄ

Najważniejsze informacje



Narzędzia AI stają się podstawowym źródłem wiedzy (i umiarkowanej prawdy) dla większości ludzi. Trendy popularnych wyszukiwarek maleją na korzyść wyszukiwań wykonywanych w chatbotach. Weryfikujemy obrazy, tekst, źródła... no i niestety czasem tworzymy też własne wersje „prawdy”.

Reakcje internautów po tragicznych wydarzeniach w Krakowie budzą wiele wątpliwości, czy chatbot jest właściwym narzędziem do weryfikacji faktów. Jednym z najczęściej powtarzanych błędów były próby poprawy jakości obrazów, aby uczynić elementy bardziej czytelnymi. Dobrym przykładem jest nadruk na koszulce, który doczekał się wielu różnych wersji.

Przeglądając treści w Internecie, trzeba mieć się na baczności, a wszelkie wysoce wyraziste grafiki powinny budzić podejrzenia. Każda grafika, która została poddana modyfikacjom, mogła utracić swój rzeczywisty przekaz, a kolejne zmiany mogą coraz bardziej zniekształcać rzeczywistość. Wykorzystują przy tym nieprawdziwą symbolikę, która potem może być niewłaściwie interpretowana w przestrzeni publicznej.



Najważniejszy wniosek:
ostry, czytelny obraz nie musi pokazywać prawdziwych szczegółów. Generatywna obróbka dopisuje to, czego nie widać w oryginale.

Co może się zmienić?

Napisy i symbole na ubraniu, twarze, przedmioty, oświetlenie, widok za oknem, a nawet pora dnia.

Dlaczego to groźne?

Wymyślony detal może być traktowany jak dowód. Każde kolejne udostępnienie utrwala fałszywą wersję wydarzeń.

Co sprawdzać?

Pierwotny plik i źródło publikacji, wcześniejsze wersje kadru, różnice w szczegółach, oficjalne komunikaty i niezależne ustalenia organizacji fact-checkerskich.

Czego nie robić?

Nie wyciągać wniosków o ludziach, symbolach ani przedmiotach z materiałów „poprawionych” przez AI. Nie udostępniać przeróbek dalej bez kontekstu.

Ale są też narzędzia dedykowane

Klasyczne wyostrzanie lub korekta jasności nie zawsze zmieniają sens obrazu. Inaczej jest z generatywną rekonstrukcją, która może wytworzyć nowe szczegóły. Z samych dwóch wersji zdjęcia nie zawsze można ustalić, jakiego narzędzia użyto ani w jakiej kolejności powstały pliki.

Kiedy AI „poprawia” zdjęcie, może też dopisać historię

Po tragicznych wydarzeniach zdjęcia i nagrania trafiają do internetu szybciej, niż można zweryfikować wszystkie towarzyszące im twierdzenia. W ostatnich dniach obok oryginalnych, słabej jakości kadrów (m.in skompresowanych przez informatyczne systemy na social media) pojawiały się ich wersje wysokiej rozdzielczości z m.in. wyraźniejszymi napisami, naszywkami i przedmiotami. Problem w tym, że część tych szczegółów nie została odkryta, tylko wygenerowana.

W kontekście ataku w Krakowie z 7 listopada 2026 r. nie jest to wyłącznie kwestia estetyki. Wymyślony znak na ubraniu czy nieprawdziwy emblemat mogą stać się punktem wyjścia do kolejnych spekulacji. Dlatego poniższe przykłady pokazują mechanizm, który utrudnia odróżnianie obrazu źródłowego od jego przetworzonej interpretacji.

Dlaczego wyraźniejsze nie znaczy wierniejsze?

Jeżeli oryginalny materiał ma niską rozdzielczość, silną kompresję lub zawiera rozmyte fragmenty, brakujących informacji nie da się po prostu „odzyskać” z istniejących pikseli. Generatywne systemy AI mogą jedynie oszacować, jakie szczegóły mogły znajdować się w danym miejscu, uzupełniając obraz elementami wyglądającymi wiarygodnie, takimi jak litery, kontury, tekstury czy fragmenty tła.

Co więcej, wynik działania takiego systemu może być podatny na sugestie użytkownika. Gdy osoba analizująca materiał jest przekonana, że przedstawia on określony symbol, napis lub zdarzenie, model może wygenerować szczegóły zgodne z tą interpretacją, nawet jeśli nie znajdują one potwierdzenia w materiale źródłowym.

Takie działania dają wizualnie przekonujący wynik, ale nie potwierdza on tego, co tak naprawdę znajdowało się przed kamerą. Mało tego, zależy to również od wykorzystanego modelu i jego danych treningowych. Na przykład na „poprawianych” zdjęciach często pojawiał się szeroki rzep na rękawie. Model nie odczytał go z oryginału. Dodał go, bo na zdjęciach, z których się uczył, zielone kurtki wojskowe zwykle mają taki element.

Istotne jest również to, że fałszywy obraz można wielokrotnie przekształcać. Każda kolejna wersja może wyglądać bardziej realistycznie i jednocześnie jeszcze bardziej odbiegać od materiału, który posłużył za punkt wyjścia.

1. Koszulka: różne nadruki w rozpowszechnianych wersjach

W internecie pojawiło się wiele wersji nadruku na koszulce. Dwie najczęściej były rozważane. Na jednej koszulka ma duży biało-zielony, geometryczny wzór, na drugiej widoczny jest rozbudowany, żółto-zielony nadruk. Różnią się również fragmenty otoczenia. Nie są to dwa identyczne kadry różniące się wyłącznie ostrością. Samo zestawienie pokazuje rozbieżności w rozpowszechnianych obrazach, cechach twarzy i proporcjach, jednak bez materiałów pierwotnych nie pozwala ustalić pochodzenia każdej wersji ani konkretnego sposobu jej wytworzenia.



FOT. 1. Porównanie trzech obrazów: kadr z nagrania zatrzymania oraz dwie rozpowszechniane wersje fotografii po zwiększeniu jakości za pomocą modeli generatywnych

Tymczasem prawdziwa koszulka wyglądała inaczej. Wirtualna Polska opublikowała zdjęcie zabezpieczonej odzieży - czarny T-shirt z biało-zielonym nadrukiem i napisem „uciekłem z morawicy”. Konkret24 potwierdził, że AI, zwiększając rozdzielczość, zmieniło grafikę na koszulce.



FOT. 2. Fotografia odzieży; źródło: Wirtualna Polska.

2. To samo okno, inna pora dnia

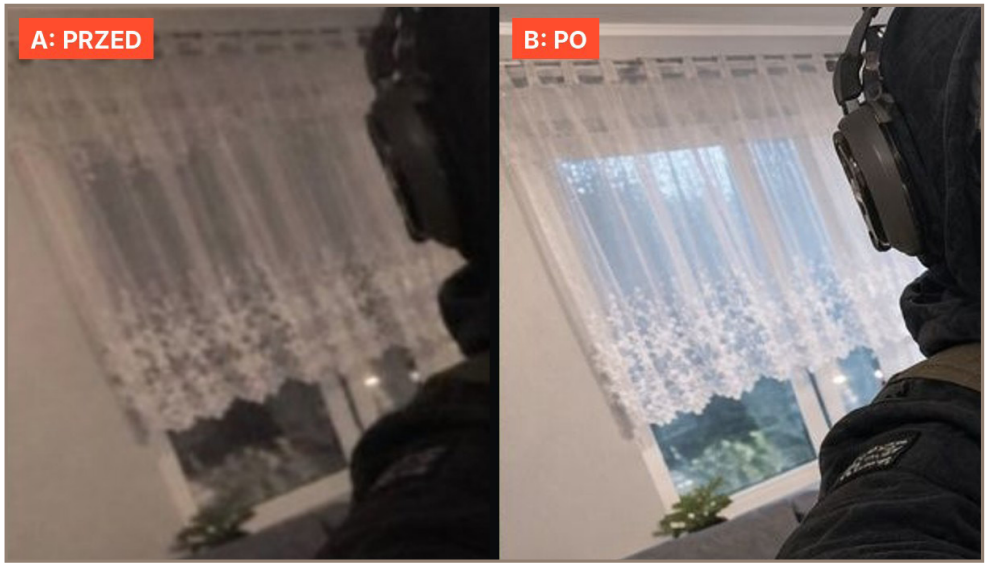
Drugi przykład pokazuje, że „poprawa” dotyczy też obrazów, które od początku były fałszywe. W sieci krążyły dwie wersje tego samego zdjęcia postaci w masce, w mieszkaniu. Obie są fałszywe. Jedna to stan przed „poprawą jakości”, druga po niej.

Na pierwszej wersji pokój jest ciemny, a za oknem panuje noc. Obraz jest słabej jakości i zaszumiony, a jedyne światło pada z lampy pod sufitem. Wygląda „autentycznie”, a mimo to już jest fałszywy. Na drugiej, za tym samym oknem jest jasny dzień i zielone drzewa. Do tego światło jest równe w całym pokoju, a na kamizelce pojawiły się nowe detale.



FOT. 3. Dwa warianty obrazu: różnice w oświetleniu i widoku za oknem.

Najlepiej widać to na oknie. Ta sama szyba, a za nią dwie pory dnia. Model nie wiedział, co jest za oknem, więc wymyślił krajobraz pasujący do „ładnego” zdjęcia.



FOT. 4. Dwa warianty obrazu: różnice w oświetleniu i widoku za oknem.

„Poprawa” nie przybliżyła fałszywki do prawdy, tylko zrobiła z niej jeszcze większą fałszywkę, z inną porą dnia. Z jednego fejka robią się dwa, potem pięć, a każda kolejna wersja wygląda wiarygodniej od poprzedniej.

3. Naszywki, numery, oznaczenia - dodawana symbolika

Na jednym z kadrów leży postać w mundurze. Na rękawie i kasku widać jasne plamy, z których nie da się odczytać kształtu, cyfr ani znaków. Po „wyostrzeniu” plamy zamieniły się w czytelną naszywkę z numerem „22”, flagę i orła na kasku.

Wystarczy porównać kształty. W oryginale naszywka jest okrągła, a w wersji AI kwadratowa. Biały kleks na kasku stał się orłem.



FOT. 5. Porównanie oryginalnych rozmyć z wersją zawierającą czytelne symbole, prezentujące przykład zmiany oznaczeń na rękawie i kasku.

Ten sam mechanizm zadziałał przy innym kadrze. Jeden z licznych wpisów z rozmazaną okrągłą naszywką miał ok. 1 mln wyświetleń. W odpowiedzi ktoś „wyostrzył” ją w AI i z plamy powstał emblemat „6”. Platforma X sama oznaczyła ten obraz jako utworzony za pomocą AI, a mimo to zaczęły się pytania, jaką formację albo grupę oznacza.



FOT. 6. Porównanie oryginalnych rozmyć z wersją zawierającą czytelne symbole, prezentujące przykład zmiany oznaczeń na rękawie i kasku.

Symbol to narracja. Na podstawie takich detali ludzie budują dodatkoweteorie związane tragedią. Doszukiwanie się ukrytej symboliki prowadzi do nieuzasadnionych spekulacji i odwraca uwagę od faktów ustalanych w toku śledztwa.

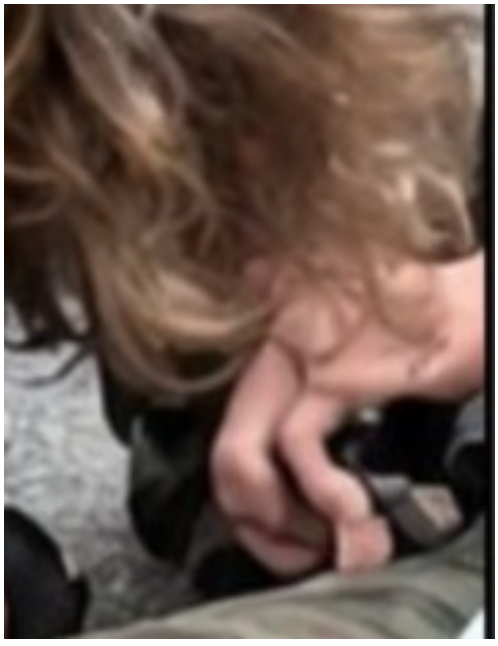
4. Broń: ta sama a jednak inna

W kolejnym zestawieniu przedmiot widoczny na zdjęciu o niskiej jakości po przetworzeniu przez AI zyskuje wyraźne detale, przy czym jego konstrukcja ulega zmianie. Oba zjawiska opierają się na podobnym mechanizmie, gdzie model generatywny uzupełnia brakujące informacje, tworząc wiarygodnie wyglądający wynik, który niekoniecznie odpowiada rzeczywistości.



FOT. 7. Zestawienie obrazu źródłowego i wersji po przetworzeniu.

Na tym samym zdjęciu w innej części kadru można znaleźć inne deformacje, np. dłoni:



FOT. 8. Deformacje dłoni spowodowane modyfikacjami AI.

5. Podmieniona twarz, nowe tło

Kolejny przykład pokazuje, jak generatywna sztuczna inteligencja zmienia możliwości manipulowania fotografiami. Punktem wyjścia było autentyczne zdjęcie z niedawnego wydarzenia publicznego, przedstawiające uścisk dłoni dwóch osób. Na jego podstawie przygotowano zmanipulowaną wersję, w której podmieniono twarz jednej z nich.

Jeszcze do niedawna tego typu manipulacje (ang. face swap) polegały przede wszystkim na zastąpieniu twarzy jednej osoby twarzą innej, z zachowaniem pozostałych elementów fotografii. Obecnie modele generatywne pozwalają

(i często to robią) na znacznie szerszą ingerencję. Po przekazaniu obrazów referencyjnych i odpowiedniej instrukcji tekstowej (image-to-image z promptem), model podmieni twarz, ale może również przebudować inne fragmenty zdjęcia. W analizowanym przykładzie zmieniły się szczegóły tła, elementy otoczenia, cechy twarzy osób w tle i jakość odwzorowania obrazu. W efekcie powstała nowa, wiarygodnie wyglądająca wersja fotografii, która może zostać odebrana jako autentyczna dokumentacja wydarzenia.



FOT. 9. Wersja zmanipulowana: podmieniona twarz i przebudowane szczegóły otoczenia.



FOT.10. Fotografia wyjściowa z wydarzenia publicznego, udostępniona do porównania. Przykład rozpowszechniony we wpisie na X: https://x.com/PiotrekzLublina/status/2119658173606907919?s=20.

AI potrafi też dopisać błędne wyjaśnienie

Coraz częściej szukając odpowiedzi lub wyjaśnień zadajemy pytania różnego rodzaju chatbotom. Należy przy tym pamiętać, że ich odpowiedzi nie mogą być bezkrytycznie uznawane za prawdę, która nie wymaga weryfikacji w niezależnych źródłach. Często głównym celem chatbotów jest udzielenie nam odpowiedzi, niezależnie od tego czy posiadają wiedzę na dany temat. Skutkować to może halucynacjami, które przekazywane dalej jeszcze bardziej poszerzają zakres dezinformacji.

Poniżej znajduje się przykład, w którym chatbot zapytany o jeden z przerobionych obrazów z wydarzeń wskazał z dużą pewnością, że przedstawia on scenę z filmu „Ciche miejsce 2”, podając dodatkowe szczegóły. Odpowiedź okazała się nieprawdziwa. Podobnie jak obraz zmodyfikowany przez AI może zawierać nieistniejące szczegóły, tak również interpretacja wygenerowana przez model językowy może sprawiać wrażenie wiarygodnej mimo braku oparcia w faktach.

Weryfikacja informacji wymaga porównania ich z niezależnymi i wiarygodnymi źródłami. Odpowiedź wygenerowana przez model językowy nie stanowi dowodu autentyczności zdjęcia ani potwierdzenia prawdziwości przedstawionej interpretacji. Może być jedynie punktem wyjścia do dalszej analizy, a nie jej ostatecznym wynikiem.

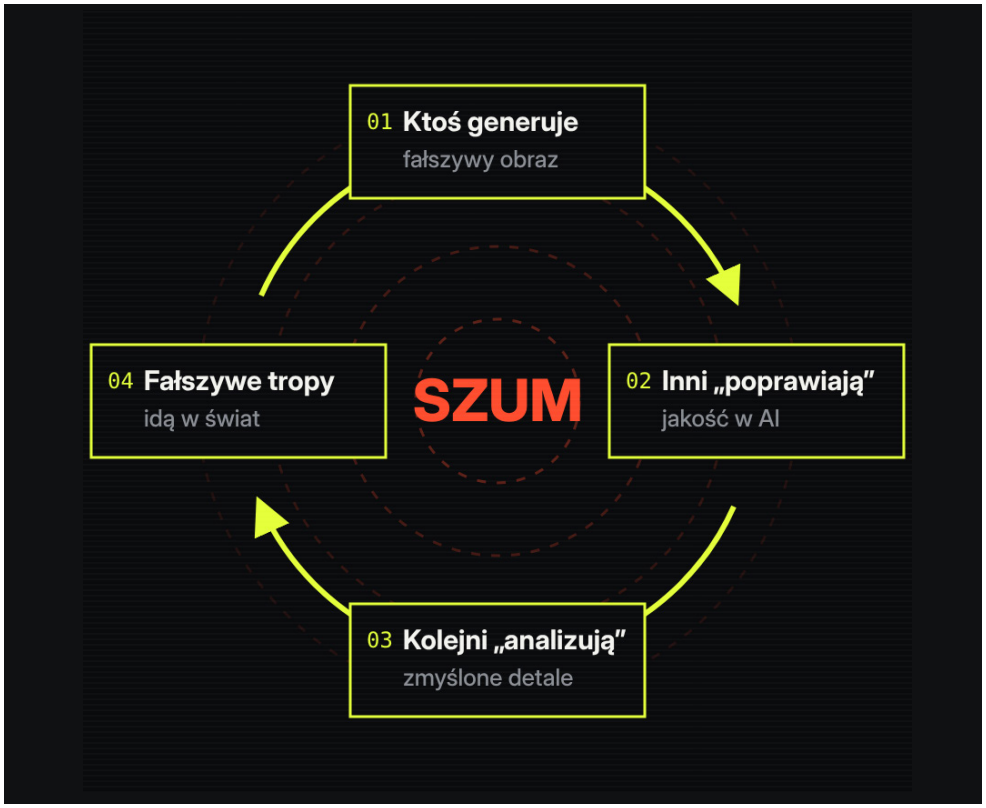


FOT.11. Przykład halucynacji chatbota w odpowiedzi na pytanie dotyczące przerobionego zdjęcia.

Spirala dezinformacji

Mechanizmy opisane powyżej prowadzą do powstania spirali dezinformacji. Świadek zdarzenia publikuje zdjęcie lub nagranie, którego jakość nie umożliwia dokładnej weryfikacji szczegółów. Następnie ktoś chcąc lepiej zrozumieć zawartość materiału, wykorzystuje narzędzia sztucznej inteligencji do poprawy jego jakości i publikuje zmodyfikowaną wersję wraz z własną interpretacją, traktując ją jako potwierdzenie określonej tezy.

Nie musi być to działanie celowe. Często wynika wyłącznie z ciekawości i chęci zobaczenia danego kadru wyraźniej. Problem narasta wtedy, gdy kolejne osoby próbują wydobyć jeszcze więcej szczegółów, ponownie przetwarzając już zmodyfikowany materiał lub korzystając z innych narzędzi AI na materiale źródłowym. Równolegle mogą pojawiać się osoby świadomie tworzące wersje dostosowane do określonej narracji, które następnie są dalej rozpowszechniane i modyfikowane.



FOT.12. Schemat mechanizmu.

Jak weryfikować zdjęcia i nagrania

1. **Zidentyfikuj źródło materiału**
   szukaj najwcześniejszej dostępnej publikacji, która zawiera dane zdjęcie lub nagranie. Zwracaj uwagę, czy prezentowany jest pełny kadr, a nie jedynie zrzut lub nagranie ekranu.

2. **Porównuj dostępne wersje materiału.**
   Analizuj, czy w kolejnych publikacjach pojawiają się nowe napisy, symbole, twarze, elementy tła lub zmiany oświetlenia. Szczególną ostrożność zachowaj w przypadku szczegółów, które w oryginalnym materiale były niewidoczne lub nieczytelne.

3. **Oddzielaj obserwację od interpretacji.**
   Rozpoznanie symbolu przez internautę lub odpowiedź wygenerowana przez chatbota nie stanowią niezależnego dowodu prezentowanych tez.

4. **Zatrzymaj rozpowszechnianie niezweryfikowanych treści.**
   Nie publikuj materiałów poddanych „ulepszeniom” AI jako dokumentacji rzeczywistych zdarzeń. Nie powielaj również treści ani twierdzeń, co do których nie masz uzasadnionej pewności, że są prawdziwe.



AI może zwiększyć czytelność obrazu, ale może jednocześnie sztucznie zawyżać pewność odbiorców co do szczegółów, których nigdy nie zarejestrowano. **W sytuacjach kryzysowych ostrożność i rzetelna weryfikacja informacji są często ważniejsze niż szybkość ich publikowania.**

Źródła

1 Demagog – „Atak w Krakowie. Uwaga na przerobione i «wyostrzone» obrazy AI”

2 Konkret24 / TVN24 – „Strzelanina w Krakowie – fejkowe zdjęcia i filmy. Nie dajcie się nabrać”

3 Wirtualna Polska – „Taśką koszulkę miał na sobie zamachowiec”

4 Rzeczpospolita – „Zamach w Krakowie. Pięć fałszywych narracji, które pojawiły się w internecie”.

Dodatkowa literatura

1 Norman i in., „An Investigation into the Impact of AI-Powered Image Enhancement on Forensic Facial Recognition”, CVPR Workshops (2024).

2 Mazzei i in., „AI-based face super resolution for forensic evaluation: a trustworthiness analysis”, DFRWS EU (2026).



NASK

Ośrodek Analizy Dezinformacji
NASK

PROJEKT FINANSOWANY ZE ŚRODKÓW MINISTERSTWA CYFRYZACJI

CWZ
CENTRUM
WALKI
Z DEZINFORMACJĄ