ANALIZA WIARYGODNOŚCI ZDJĘĆ I FILMÓW Z WYKORZYSTANIEM OSINT

CREDIBILITY ANALYSIS OF PHOTOS AND VIDEOS USING OSINT

dr inż. KRZYSZTOF WOSIŃSKI

TELDAT sp. z o.o. sp.k.

Streszczenie

W dobie coraz większych możliwości takiej manipulacji obrazami oraz materiałami wideo (głównie przy wykorzystaniu mechanizmów AI), aby nie było to możliwe do wychwycenia przez przeciętnego użytkownika portali internetowych i mediów społecznościowych, koniecznym jest opracowanie sposobów prostej analizy treści wizualnych w Internecie. W niniejszym artykule autor opisuje rodzaje zniekształceń informacji, podejście oparte na zasadach rozpoznania otwartoźródłowego do ich analizy, a także przedstawia wybrane techniki i narzędzia z tego zakresu, które są w stanie pomóc odbiorcom w prostym i szybkim wychwyceniu prawdopodobnych manipulacji w materiałach graficznych.

Słowa kluczowe: dezinformacja, weryfikacja, OSINT, rozpoznanie otwartoźródłowe, media społecznościowe, cyberbezpieczeństwo.

Summary

In the era of increasing opportunities to manipulate image and video materials (mainly by using AI mechanisms) to such an extent that it becomes impossible for the average user of internet portals and social media to discern it, it is necessary to develop simple methods for analysis of the visual content on the internet. In this article, the author describes the types of information distortion, the approach to information analysis based on open-source intelligence principles, and presents selected techniques in this matter which offer simple and fast methods of detecting probable manipulations in the graphical content.

Keywords: disinformation, verification, OSINT, open-source intelligence, social media, cybersecurity.

Wprowadzenie

Każda informacja, która jest pozyskiwana ze źródeł internetowych, musi zostać zweryfikowana, aby mogła stanowić podstawę do podejmowania decyzji oraz formułowania wniosków. Bez tego kroku wszelkie podjęte decyzje będą podatne na działania zaburzające wiarygodność informacji, a co za tym idzie – obraz rzeczywistego stanu wydarzeń i faktów. Jednym z rodzajów technik mogących pomóc w weryfikacji informacji są techniki rozpoznania wykonywanego w oparciu o otwarte źródła (ang. open-source intelligence – OSINT). Poznanie podstawowych technik manipulacji pozwala zweryfikować prawdziwość przekazu poprzez porównanie go z innymi źródłami dostępnymi w Internecie, a także zidentyfikować możliwe przekłamania. Łatwość dostępu do Internetu i do możliwości pobierania zawartości stron (np. obrazów lub filmów) oraz publikowania treści w sieci, szczególnie w mediach społecznościowych, które stanowią jedno ze źródeł pozyskiwania informacji o aktualnych wydarzeniach, powoduje, że zagrożenie wynikające z manipulacji informacjami jest tematem wymagającym coraz większej uwagi.

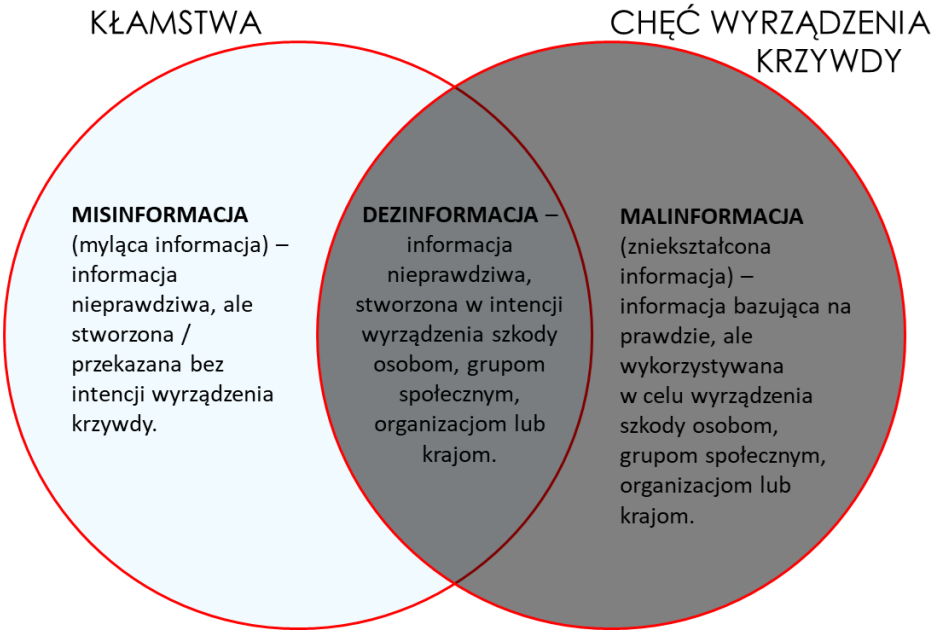
Celem niniejszej pracy było zatem zebranie i przedstawienie najważniejszych sposobów przeciwdziałania dezinformacji, wynikającej ze zmanipulowania materiałów źródłowych (celowo lub nienaumyślnie). Analiza problemu dotyczyła w szczególności weryfikacji informacji publikowanych w Internecie: w jaki sposób możliwe jest zidentyfikowanie manipulacji materiałów wizualnych z wykorzystaniem rozpoznania otwartoźródłowego oraz jakie narzędzia umożliwiają sprawną analizę prawdziwości przekazów?

Artykuł poprzez zgłębienie możliwości weryfikacji przekazów za pomocą ich porównywania z innymi wyszukanymi informacjami z użyciem technik OSINT-owych wykazuje, iż kluczowa jest umiejętność szybkiego rozpoznawania zmanipulowanych informacji przez każdego odbiorcę w Internecie, bez której będzie on podatny na dezinformację.

Rodzaje zaburzonych informacji

Informacje, które zostały w pewnym stopniu zaburzone lub zniekształcone (celowo lub nie), można podzielić na dwie kategorie: dezinformację oraz misinformację. Wydany przez UNESCO poradnik dla dziennikarzy „Journalism, ‘Fake News’, and Disinformation” terminem „dezinformacja” określa informacje, które są nieprawdziwe, a osoby, które je rozpowszechniają, robią to świadomie w celu wyrządzenia szkody. Jako definicję „misinformacji” poradnik ten z kolei wskazuje informacje, które są nieprawdziwe, ale osoby je rozpowszechniające nie są świadome tego faktu. Wskazane jest tam także inne określenie – „mal-informacja”, czyli informacja, która jest oparta na faktach, jednak jest wykorzystywana w celu wyrządzenia szkody osobom, organizacjom

lub krajom[1]. Możliwe jest więc stworzenie wizualnej reprezentacji powiązań pomiędzy wskazanymi określeniami, którą przedstawiono na poniższym rysunku.



Rysunek 1 – Podobieństwa i różnice pomiędzy misinformacją, malinformacją a dezinformacją. Źródło: opracowanie własne autora.

Niekiedy można spotkać się także z określeniem „fake news" w odniesieniu do fałszywych informacji, jednak samo użycie określenia „news" (z ang. wiadomości) stoi w sprzeczności z opisywanymi tutaj zmanipulowanymi informacjami, gdyż powinno ono określać jedynie zweryfikowane informacje, przekazywane w interesie publicznym. Określenie to może być co najwyżej traktowane jako oksymoron, który jednak stawia w niekorzystnym świetle wiarygodne wiadomości poprzez stwarzanie niekorzystnych skojarzeń ze słowem „news" u odbiorców[2].

Niebezpieczeństwa wynikające z użycia dezinformacji wymuszają stosowanie jak największego wachlarza technik, dostępnych dla jak najszerszego grona odbiorców informacji, w celu jak najłatwiejszej jej weryfikacji. W kolejnych podrozdziałach niniejszego artykułu przedstawione zostały podstawowe techniki rozpoznania otwartoźródłowego, które pozwalają na prostą analizę materiałów w celu zidentyfikowania potencjalnych zafałszowań informacji. Dla zaawansowanej analizy

[1] Ireton, Cheilyn; Posetti, Julie, *Journalism, 'Fake News', and Disinformation*, UNESCO, Paryż, 2018.
[2] Ibidem.

bardziej złożonych kampanii dezinformacyjnych i innych działań przestępczych wskazane jest używanie kombinacji wielu technik i użycia bardziej zaawansowanych narzędzi.

Podstawowe zasady wykorzystania rozpoznania otwartoźródłowego do analizy informacji

Rozpoznanie otwartoźródłowe, określane także często angielskim terminem skrótowym OSINT, oznacza działania wywiadowcze, prowadzone na bazie publicznie dostępnych informacji[3], co pozwala na wykorzystanie technik z nimi związanych niemal przez każdą osobę, która posiada dostęp do Internetu i zapoznana jest z podstawowymi sposobami analizy danych. Sam dostęp do otwartych źródeł w tym przypadku nie jest wystarczający, ponieważ aby poprawnie zweryfikować przekaz, należy mieć na uwadze możliwe błędy popełniane w trakcie tych działań, a także metody przeciwdziałania nim[4]. W celu stworzenia odpowiedniego nastawienia osób analizujących informacje wskazane jest korzystanie z tzw. zasady ABC: assume nothing, believe no one, check everything (z ang. „nie bierz niczego za pewnik, nie wierz nikomu, wszystko sprawdzaj").

Jedną z najprostszych technik wykorzystania OSINT-u do weryfikacji informacji jest porównanie przekazów pochodzących z różnych źródeł. Przykładem może być tutaj skorzystanie z różnych narzędzi mapowych w celu sprawdzenia informacji o zatorach na drodze, jak to miało miejsce w pierwszym dniu ataku wojsk rosyjskich na terytorium Ukrainy 24 lutego 2022 roku, kiedy to na rosyjskich mapach Yandex nie były widoczne żadne utrudnienia na drodze prowadzącej z Biełgorodu do ukraińskiej granicy, a mapy Google wskazywały znaczne spowolnienie ruchu[5]. Niezależność jednego ze źródeł w tym przypadku była jednym z pierwszych sygnałów, że wojska rosyjskie zaczynają przemieszczać się w kierunku granicy z Ukrainą i powodują korki na wskazanej drodze.

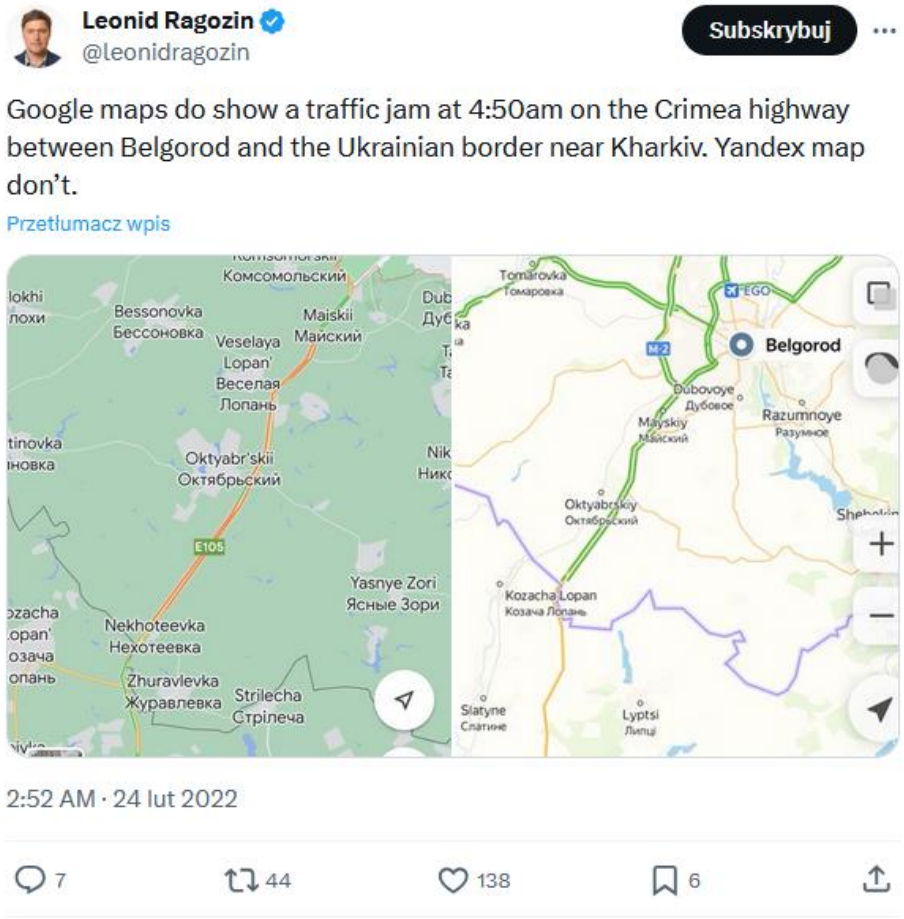
Podobnie porównanie wyników wyszukiwarek tekstowych może doprowadzić do rozszerzenia świadomości o źródłach przekazujących alternatywne informacje na poszukiwany temat. Ponad 80% użytkowników Internetu na świecie korzysta z wyszukiwarki Google, podczas gdy na każdą z pozostałych wyszukiwarek, takich jak Bing, Yandex, Yahoo czy Baidu nie przypada więcej niż niecałe 10 procent

[3]The Interagency OPSEC Support Staff, *Operations Security INTELLIGENCE THREAT HANDBOOK*, 1996, https://irp.fas.org/nsa/ioss/threat96/index.html. Dostęp: 19.11.2024.

[4] Wosiński, Krzysztof, *Bezpieczeństwo osób i systemów IT z wykorzystaniem białego wywiadu. OSINT*, PWN, Warszawa, 2024.

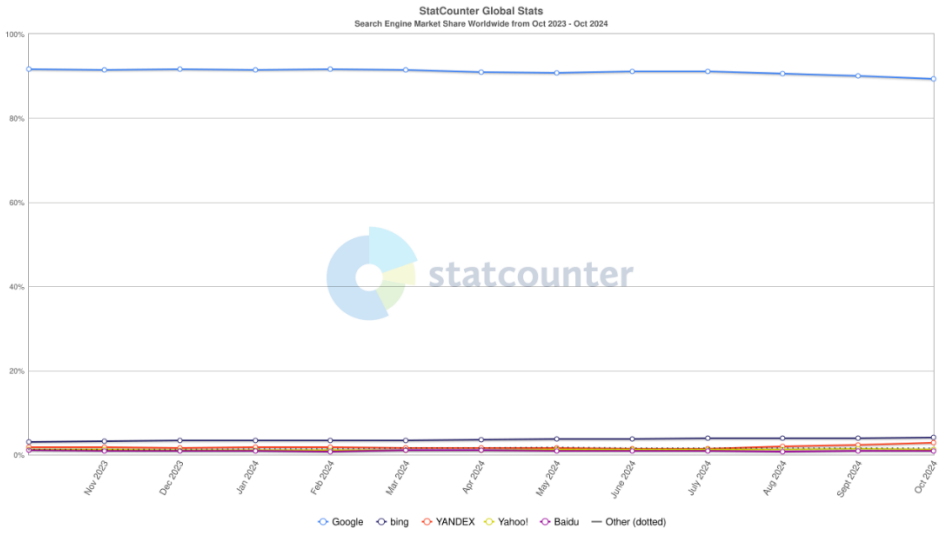
[5]Ragozin, Leonid [@leonidragozin], „Google maps do show a traffic jam at 4:50am on the Crimea highway between Belgorod and the Ukrainian border near Kharkiv. Yandex maps don't.", *X*, 24 lutego 2024, https://x.com/leonidragozin/status/1496664359761682433.

użytkowników[6], a według niektórych źródeł nawet mniej niż 5 procent[7]. Możliwe jest więc ukrycie (celowe lub przypadkowe) pewnych faktów i kontrstanowisk przed odbiorcami. Dlatego skorzystanie z wielu źródeł otwartych danych (np. właśnie wyszukiwarek), jest wskazane w celu wykonania poprawnej analizy docierającej do odbiorcy informacji.



Rysunek 2 – Wpis w portalu X wskazujący na różnice w informacjach przekazywanych na mapach Google i Yandex w dniu ataku Rosji na Ukrainę. Źródło: x.com

[6]*Market share of leading desktop search engines worldwide from January 2015 to January 2024*, https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/. Dostęp: 17.11.2024.
[7]*Search Engine Market Share Worldwide, Oct 2023 - Oct 2024*, https://gs.statcounter.com/search-engine-market-share. Dostęp: 17.11.2024.



Rysunek 3 – Widoczna różnica pomiędzy liczbą użytkowników wyszukiwarki Google i pozostałych wyszukiwarek wg portalu statcounter Źródło: statcounter.com

Poprzez wyszukiwanie podobnych treści jak w przekazie, co do którego prawdziwości odbiorca ma wątpliwości, możliwe jest również zidentyfikowanie obrazów, które pochodzą z innych źródeł, a zostały użyte w danym kontekście, w celu uprawdopodobnienia zmanipulowanego przekazu.

Przykład analizy dezinformacji wykorzystującej materiały wizualne

W celu identyfikacji możliwości wykorzystania obrazów lub materiałów wideo pochodzących z wcześniejszych przekazów, a wykorzystanych do zafałszowania informacji, możliwe jest skorzystanie z wyszukiwarek obrazowych (np. Google Grafika, wyszukiwanie wizualne Bing, YandexImages lub TinEye), które wskażą, czy dany obraz nie występuje w innych źródłach internetowych. W przypadku znalezienia obrazów na stronach, które oznaczone są datami wpisów, można łatwo stwierdzić, czy dany obraz lub materiał wideo związany jest np. z bieżącymi wydarzeniami, czy też został zarejestrowany już jakiś czas temu, w związku z innym wydarzeniem.

Przykładem prostego śledztwa, które jednak nie ustrzegło się błędu, jest analiza pochodzenia zdjęcia ilustrującego wpis z 23 grudnia 2023 na portalu X, dotyczący rzekomej przemocy żołnierzy izraelskich na palestyńskiej dziewczynie[8]. Autor artykułu, w którym opisane zostało wspomniane śledztwo, wskazuje na wyszukiwanie obrazu

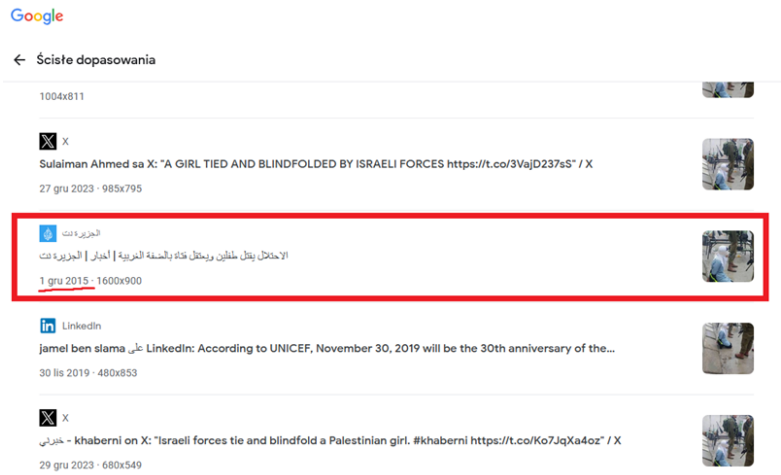
[8]Armengol, Joaquim, *Decoding disinformation. The power of open source intelligence*, The Natolin Blog, College of Europe in Natolin, https://natolinblog.eu/2024/03/decoding-disinformation-the-power-of-open-source-intelligence/. Dostęp: 17.11.2024.

pobranego z wpisu na portalu X i analizę wyników w celu odnalezienia jego wcześniejszych wystąpień. O ile postawa taka jest godna pochwały i pokazuje możliwości dostępnej praktycznie dla każdego analizy zafałszowanego przekazu, to jednocześnie dowodzi także konieczności bardzo dokładanego zbadania uzyskanych wyników wyszukiwania.

Rysunek 4 – Wpis z 26.12.2023 roku mówiący o zatrzymaniu palestyńskiej dziewczyny. Źródło: x.com

Autor odnalazł wcześniejszy artykuł, zilustrowany tym samym zdjęciem, dotyczący zatrzymania palestyńskiej dziewczyny we Wschodniej Jerozolimie, jednak błędnie założył on, że zgodnie z datą opublikowania artykułu miało to miejsce w 2017 roku. W artykule tym jednak zdjęcie jest bardzo niskiej jakości i oznaczone jako pochodzące z mediów społecznościowych, bez wskazania konkretnego źródła. Może to budzić pewne wątpliwości co do jego wiarygodności. Autor śledztwa nie odnalazł jednak jeszcze wcześniejszego artykułu, pochodzącego z grudnia 2015 roku i opublikowanego na stronie Aljazeera, w którym widnieje to samo zdjęcie, jednak w wysokiej rozdzielczości, a jego podpis wskazuje na aresztowanie palestyńskiej dziewczyny pod pretekstem posiadania noża. O ile prawidłowo zostało wskazane, że zdjęcie zamieszczono w innym kontekście i dotyczyło ono dużo wcześniejszego zdarzenia, to przykład ten pokazuje, że niekiedy łańcuch kolejnych momentów wykorzystania obrazu może być bardzo długi i zawsze

należy dążyć do uzyskania jak najwcześniejszego wyniku, który może okazać się tym pierwotnym. Często będzie miał on też najlepszą jakość materiału, gdyż dużo łatwiej jest zmanipulować przekaz, jeżeli niektóre elementy obrazu będą trudno dostrzegalne, a co za tym idzie, trudniej będzie dociec, czy zdjęcie to (lub zapis wideo) dotyczy faktycznie opisywanego wydarzenia, czy nie.



Rysunek 5 – Wynik wyszukiwania z datą publikacji 01.12.2015. Źródło: google.com

Rysunek 6 – Artykuł z 01.12.2015 ze zdjęciem, które później było używane w innych artykułach, również w innym kontekście. Źródło: aljazeera.net[9]

[9] الاحتلاليقتلطفلينويعتقلفتاةبالضفةالغربية [Okupanci zabijają dwójkę dzieci i aresztują dziewczynę na Zachodnim Brzegu], Al. Jazeera,

Weryfikacja prawdziwości materiałów wizualnych powinna uwzględniać 5 filarów[10]. Są to sprawdzenia dotyczące:

1. pochodzenia – czy mamy do czynienia z treścią źródłową?
2. źródła – kto stworzył lub uwiecznił analizowane treści?
3. czasu – kiedy stworzono analizowane treści?
4. miejsca – z jaki miejscem są związane analizowane treści / skąd pochodzi konto je rozpowszechniające?
5. motywacji – dlaczego powstała strona lub konto publikujące analizowane treści albo w jakich okolicznościach uchwycono dane zdarzenie?

Zadanie sobie powyższych pytań może pomóc zweryfikować częste próby zafałszowania informacji poprzez posiłkowanie się obrazami, które:

- zostały wykonane w innym miejscu niż twierdzi przekaz;
- zostały wykonane w innym czasie niż twierdzi przekaz;
- zostały użyte w innym kontekście niż oryginalny przekaz.

Nie zawsze udaje się uzyskać odpowiedź na postawione powyżej pytania, szczególnie w zakresie motywacji stojącej za utworzeniem danego przekazu, jednak próba zweryfikowania informacji w tychże zakresach może znacznie usprawnić proces zdemaskowania fałszywego przekazu. Już sama weryfikacja innych treści publikowanych w Internecie przez analizowany serwis lub konto w mediach społecznościowych umożliwia określenie poziomu prawdopodobieństwa, że dany materiał może być prawdziwy lub nie. Przykładowo, jeśli konto publikuje wyłącznie treści silnie nacechowane emocjonalnie, można podejrzewać, że próbuje ono wpłynąć na ocenę sytuacji przez jego czytelnika.

Wykorzystanie narzędzi OSINT-owych do analizy danych

Do prostych, a jednocześnie dość potężnych narzędzi umożliwiających analizę obrazów należą wyszukiwarki obrazowe (ang. RIS – Reverse Image Search), które z wykorzystaniem obrazu źródłowego wskazują z jednej strony na miejsca w Internecie, w których można znaleźć ten sam obraz, a z drugiej – coraz częściej są w stanie zanalizować zawartość danego obrazu, wskazując np. na prawdopodobne miejsce wykonania danego zdjęcia lub obiekty, które się na nim znajdują.

W ramach wyszukiwarki Google Grafika dostępne jest narzędzie Google Lens, które jest w stanie bardzo dobrze zanalizować zawartość obrazu we wspomnianym wyżej zakresie.

---

https://www.aljazeera.net/news/2015/12/1/%D8%A7%D9%84%D8%A7%D8%AD%D8%AA%D9%84%D8%A7%D9%84-%D9%8A%D9%82%D8%AA%D9%84-%D8%B7%D9%81%D9%84%D9%8A%D9%86-%D9%88%D9%8A%D8%B9%D8%AA%D9%82%D9%84-%D9%81%D8%AA%D8%A7%D8%A9. Dostęp: 17.11.2024.

[10] Urbani, Shaydanay, First Draft's Essential Guide to Verifying Online Information, First Draft, październik 2019.

Umożliwia także przedstawienie listy stron internetowych, na których znajduje się badany obraz, wraz z przedstawieniem dat, w których dany zasób został opublikowany, co pozwala na łatwe odnalezienie pierwotnego miejsca jego udostępnienia w sieci.

Podobne wyszukiwarki oferują serwisy Bing i Yandex, a także coraz więcej serwisów, również sprzedażowych, które dodały do swoich wyszukiwarek tekstowych wyszukiwanie za pomocą obrazu w celu lepszego odnalezienia pożądanego przez klienta przedmiotu. Wyszukiwarki Google oraz Yandex mają również możliwość rozpoznawania tekstu zawartego na zdjęciu, przez co mogą znacznie ułatwić sprawdzenie, czy zostało ono wykonane w miejscu, na które wskazuje przekaz (np. zdjęcie z Chin posiada napisy w języku japońskim, co może oznaczać, że nie przedstawia ono opisywanego miejsca w tym kraju) lub zidentyfikowanie dodatkowych informacji o miejscu wykonania zdjęcia poprzez analizę zawartości napisów (np. szyldów sklepów, napisów na pojazdach itp.).

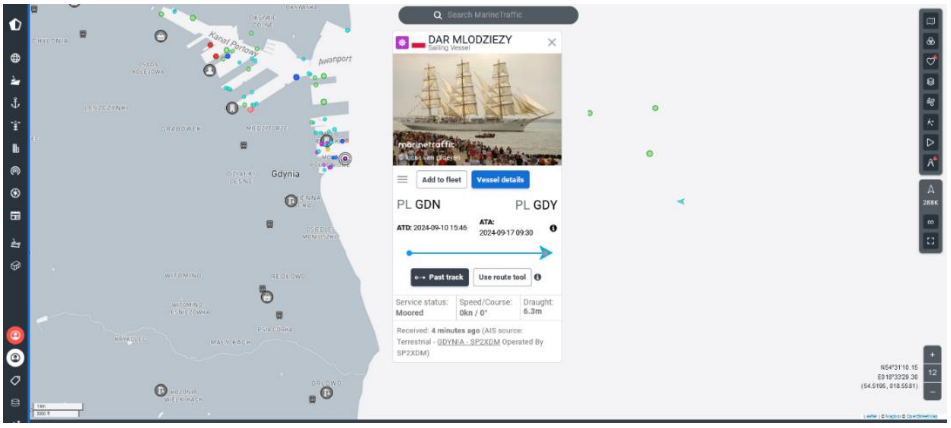
Rysunek 7 – Porównanie możliwości analizy obrazów poprzez wyszukiwarki Google, Bing i Yandex[11]

Jedną z najstarszych wyszukiwarek wykorzystujących obraz jako dane wejściowe jest serwis TinEye, który oferuje jednak jedynie dokładne wyszukiwanie innych źródeł publikacji danego obrazu, bez analizy jego zawartości, ale szczyci się dodatkowo umiejętnością odnajdywania obrazów nieznacznie zmodyfikowanych (np. zmienionych na obrazy w skali szarości).

Weryfikacja danych w różnych obszarach tematycznych możliwa jest poprzez specyficzne dla danego obszaru narzędzia wspomagające rozpoznanie otwartoźródłowe.

[11]Wosiński, Krzysztof, *Google'a szkiełko i oko. Czyli co nowego w wyszukiwaniu zawartością obrazów [czwartki z OSINTem]*, https://sekurak.pl/googlea-szkielko-i-oko-czyli-co-nowego-w-wyszukiwaniu-zawartoscia-obrazow-czwartki-z-osintem/. Dostęp: 17.11.2024.

Przykładem mogą być analizy dotyczące przemieszczania się statków, okrętów lub samolotów. Poprzez otwartość protokołów komunikacyjnych, służących do przekazywania informacji o położeniu jednostek powietrznych lub morskich, możliwe jest zweryfikowanie przekazów medialnych, które ich dotyczą. Niestety, ten fakt powoduje również, że dane takie mogą być zafałszowane i muszą zostać zweryfikowane przy użyciu innych kanałów informacyjnych. Za przykład takiego działania dezinformacyjnego może służyć tutaj informacja z 18 czerwca 2021 roku, kiedy to dwa okręty NATO, HMS Defender oraz HNLMS Evertsen, zgodnie z informacjami o ich pozycji miały być obecne dwie mile morskie od wejścia do portu w Sewastopolu. Okazało się jednak, że dane w systemie AIS (ang. Automatic Identification System) zostały zmanipulowane, a nagrania wideo z kamer portowych w Odessie oraz publikacje w mediach społecznościowych przez świadków zaprzeczyły tym doniesieniom. Otwartość takich systemów, jak wspomniany AIS czy system informowania o pozycji statków powietrznych ASD-B, powoduje, że manipulacja tymi danymi jest ryzykiem, które należy zawsze brać pod uwagę przy analizie dotyczącej tego zakresu informacji.



Rysunek 8 – Portal marinetraffic.com, pokazujący ogólnodostępne informacje o jednostkach pływających. Źródło: marinetraffic.com

Modyfikacje zdjęć w mediach społecznościowych – zagrożenia i techniki wykrywania

Oprócz manipulacji zawartością obrazów załączanych do wpisów tekstowych w mediach społecznościowych, zagrożeniem jest także modyfikacja lub wykorzystywanie kradzionych zdjęć profilowych. Przed upowszechnieniem się narzędzi wykorzystujących mechanizmy szeroko rozumianej sztucznej inteligencji do generowania obrazów, takich jak np. Midjourney, popularnym serwisem z bazą nieistniejących twarzy był thispersondoesnotexist.com. Portal ten oferował zdjęcia osób generowane przy wykorzystaniu mechanizmów GAN (generatywnych sieci współzawodniczących – ang. generative adversarial network). Obrazy tworzone przez tę technologię były jednak

bardzo łatwe do zidentyfikowania[12]. Wraz ze wzrostem możliwości wspomnianych narzędzi AI coraz trudniejsze staje się odróżnienie sztucznie wygenerowanego zdjęcia profilowego od zdjęcia rzeczywistego, jednak nie jest to niemożliwe. W wielu przypadkach nadal wspomniane narzędzia tworzą błędne obrazy, szczególnie w zakresie liczby palców, nieprawidłowości w okolicach oczu i uszu, ozdób na ciele, cieni lub napisów. Możliwe jest zniwelowanie tych niedoskonałości poprzez zachowanie dużej staranności przy generowaniu nieprawdziwego zdjęcia, jednak w wielu przypadkach operatorzy fałszywych kont nie zadają sobie tyle trudu, aby tego dokonać.

Podczas fałszowania zdjęć prawdziwych niekiedy także analiza cienia może wskazać na to, które elementy zostały zmienione, ponieważ w przypadku np. dodania lub zmiany obiektów, które są trzymane przez osoby na zdjęciu, fałszerze zapominają o konieczności zmiany również w zakresie właśnie cienia.

Rozwój technologii obecnej nawet w telefonach komórkowych umożliwia modyfikowanie rzeczywistych zdjęć z dużo większą dokładnością. W sierpniu 2024 roku na stronie magazynu The Verge ukazał się artykuł wskazujący na niebezpieczeństwa wynikające z nowej funkcji „Reimagine" w smartfonach Pixel 9 od Google[13]. Z wykorzystaniem tego narzędzia można było do zdjęć wykonanych aparatem tego smartfona dodawać elementy nierzeczywiste, takie jak zniszczone pojazdy, wypadki czy ciała. Wynik tych działań, opartych na AI, był tak dobry, że aktualnie nie ma technicznych możliwości przeciwdziałania takim zabiegom i wykrywania zmian wykonanych przez tego typu oprogramowanie. Każdy, kto ma dostęp do tego rodzaju funkcji w swoim urządzeniu mobilnym, może zastąpić dowolny element rzeczywistego obrazu stworzonymi przez mechanizmy AI obiektami. Musi jedynie umieć dobrze napisać tzw. prompta, czyli opis tego, co ma zostać wygenerowane. Poinformowane o tym fakcie przez dziennikarzy Google odpowiedziało, że tego typu działania są niezgodne z regulaminem korzystania z tego oprogramowania, jednak nie wydaje się, aby osoby tworzące dezinformacyjne obrazy miały przejąć się faktem, że łamią regulamin serwisu.

W przypadku zdjęć zmanipulowanych przy wykorzystywaniu mechanizmów AI bardzo trudno za pomocą technik wizualnych jest zidentyfikować wprowadzone zmiany, gdyż jakość generowanych obrazów jest bardzo wysoka, a dodane lub zmienione obiekty są wstawione w bardzo poprawny sposób w zakresie kolorystyki i cienia. W tym przypadku ich weryfikacja musi polegać na porównaniu informacji z innymi źródłami, opisującymi zdarzenia w tym samym czasie i miejscu, aby skonfrontować różne przekazy.

---

[12]Ibidem.

[13]Johnson, Allison, *Google's AI 'Reimagine' tool helped us add wrecks, disasters, and corpses to our photos*, The Verge, https://www.theverge.com/2024/8/21/24224084/google-pixel-9-reimagine-ai-photos. Dostęp: 17.11.2024.

Rysunek 9 – Przykład wstawienia elementów fikcyjnych na rzeczywiste zdjęcie przy pomocy mechanizmów AI. Źródło: The Verge

Sfabrykowane zdjęcia mogą być oczywiście wykorzystywane także w słusznym celu, w trakcie tworzenia kont tzw. kukiełek (ang. sock puppets) przez śledczych[14], jednak techniki te stosowane są przeważnie w zakresie działania kont o charakterze przestępczym lub dezinformacyjnym.

Standardy weryfikacji informacji

W celu ułatwienia i uwspólnienia sposobów weryfikacji informacji można posłużyć się standardami dotyczącymi ich opisu. Jednym z takich standardów jest DISARM – Disinformation TTP (Tactics, Techniques and Procedures) Framework (dawniej AMITT – Adversarial Misinformation and Influence Tactics and Techniques)[15]. Jest to ustrukturyzowane podejście, pozwalające lepiej opisywać i rozumieć zdarzenia związane z dezinformacją. Zostało ono przygotowane w taki sposób, aby wpisywało się w już istniejące podejście do opisu zagrożeń w środowisku bezpieczeństwa informatycznego. Struktura DISARM bazuje na MITRE ATT&CK, a możliwość wykorzystania sformalizowanych obiektów STIX do opisu różnego rodzaju elementów kampanii dezinformacyjnych (jak np. osoby, wzorce ataku, kampanie, powiązane grupy itp.)

[14]Wosiński, Krzysztof, *Wykorzystanie kont „alternatywnych tożsamości" w działaniach prowadzonych w Internecie*, Cybersecurity&Cybercrime, numer 2024/2 (4), Gdynia, 2024.
[15] DISARM, https://github.com/DISARMFoundation/DISARMframeworks/. Dostęp: 19.11.2024.

pozwala na łatwą wymianę tych informacji pomiędzy różnymi organizacjami zajmującymi się rozpoznawaniem i zwalczaniem tego typu działań.

Kolejnym przykładem ustandaryzowanego podejścia do weryfikacji informacji jest metoda CRAAP[16], której nazwa odnosi się do pierwszych liter angielskich słów: Currency (aktualność), Relevance (związek z tematem), Authority (tożsamość), Accuracy (dokładność), Purpose (cel). Odpowiedzi na pytania z nimi związane mogą pomóc w ocenie wiarygodności analizowanej informacji. Przykładowe pytania we wskazanych zakresach mogą być następujące:

- **Currency (aktualność):**
  - Kiedy powstała informacja?
  - Czy aktualność tej informacji jest istotna?
  - Czy przekaz był aktualizowany lub edytowany?
- **Relevance (związek z tematem):**
  - Czy informacja jest zgodna z tematyką, w ramach której jest analizowana?
  - Czy to jest jedyne źródło tej informacji?
  - Do kogo skierowany jest przekaz?
- **Authority (tożsamość):**
  - Kto stworzył / publikuje informację?
  - Czy autorzy posiadają doświadczenie, umożliwiające wypowiadanie się na dany temat?
  - Czy można dotrzeć łatwo do informacji o autorze?
  - Czy domena internetowa, w ramach której publikowana jest informacja, jest zgodna z kontekstem informacji?
- **Accuracy (dokładność):**
  - Czy informacja była przeglądana / recenzowana przed publikacją?
  - Czy informacja jest poparta faktami?
  - Czy autor wskazuje źródła?
  - Czy są obecne błędy gramatyczne lub ortograficzne?
- **Purpose (cel):**
  - W jakim celu stworzono informację?
  - Czy informacja była stworzona w celu powiadomienia, edukacji, sprzedaży, zabawy czy wpłynięcia na odbiorcę?
  - Czy intencja przekazu jest jasna?

Uzyskanie odpowiedzi na wskazane powyżej pytania może rozjaśnić cel przekazu osobie analizującej potencjalną dezinformację i ujawnić niespójności lub złe intencje publikujących daną informację.

---

[16] *Evaluating Information – Applying the CRAAP Test*, Meriam Library California State University, Chico, https://library.csuchico.edu/sites/default/files/craap-test.pdf. Dostęp: 19.11.2024.

Powyższa lista nie jest oczywiście zbiorem zamkniętym metod weryfikacji informacji w Internecie, jednak wskazuje na fakt, że do takich działań nie są potrzebne zaawansowane narzędzia ani techniki.

Podsumowanie

Rozwój Internetu i coraz szersze jego zasoby, a także coraz większe poleganie jego użytkowników na informacjach przekazywanych nie tylko na stronach zajmujących się publikowaniem zweryfikowanych wiadomości, ale głównie na przekazach pojawiających się w mediach społecznościowych, powoduje, że niezbędnym staje się umiejętność szybkiego rozpoznawania dezinformacji przez każdego odbiorcę. Działania OSINT-owe umożliwiają użytkownikom sieci proste porównanie przekazów z innymi wyszukanymi informacjami, a przez to weryfikację ich prawdziwości i motywacji stojącej za danym przekazem. Edukacja w tym zakresie wydaje się być kluczowa w celu uniemożliwienia łatwego manipulowania społeczeństwem zarówno na poziomie pojedynczych aktów przestępczych, jak i działań na arenie międzynarodowej.

Bibliografia

1. الاحتلاليقتلطفلينويعتقلفتاةبالضفةالغربية [Okupanci zabijają dwójkę dzieci i aresztują dziewczynę na Zachodnim Brzegu], Al. Jazeera, https://www.aljazeera.net/news/2015/12/1/%D8%A7%D9%84%D8%A7%D8%AD%D8%AA%D9%84%D8%A7%D9%84-%D9%8A%D9%82%D8%AA%D9%84-%D8%B7%D9%81%D9%84%D9%8A%D9%86-%D9%88%D9%8A%D8%B9%D8%AA%D9%82%D9%84-%D9%81%D8%AA%D8%A7%D8%A9. Dostęp: 17.11.2024.
2. Armengol, Joaquim, *Decoding disinformation. The power of open source intelligence*, The Natolin Blog, College of Europe in Natolin, https://natolinblog.eu/2024/03/decoding-disinformation-the-power-of-open-source-intelligence/. Dostęp: 17.11.2024.
3. DISARM, https://github.com/DISARMFoundation/DISARMframeworks/. Dostęp: 19.11.2024.
4. *Evaluating Information – Applying the CRAAP Test*, Meriam Library California State University, Chico, https://library.csuchico.edu/sites/default/files/craap-test.pdf. Dostęp: 19.11.2024.
5. Ireton, Cheilyn, Posetti, Julie, *Journalism, 'Fake News', and Disinformation*, UNESCO, Paryż, 2018.
6. Johnson, Allison, *Google's AI 'Reimagine' tool helped us add wrecks, disasters, and corpses to our photos*, The Verge, https://www.theverge.com/2024/8/21/24224084/google-pixel-9-reimagine-ai-photos. Dostęp: 17.11.2024.

7. *Market share of leading desktop search engines worldwide from January 2015 to January 2024*, https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/. Dostęp: 17.11.2024.
8. Ragozin, Leonid [@leonidragozin], „Google maps do show a traffic jam at 4:50am on the Crimea highway between Belgorod and the Ukrainian border near Kharkiv. Yandex maps don't.", *X*, 24 lutego 2024, https://x.com/leonidragozin/status/1496664359761682433.
9. *Search Engine Market Share Worldwide, Oct 2023 - Oct 2024*, https://gs.statcounter.com/search-engine-market-share. Dostęp: 17.11.2024.
10. The Interagency OPSEC Support Staff, *Operations Security INTELLIGENCE THREAT HANDBOOK*, 1996, https://irp.fas.org/nsa/ioss/threat96/index.html. Dostęp: 19.11.2024.
11. Urbani, Shaydanay, First Draft's Essential Guide to Verifying Online Information, First Draft, październik 2019.
12. Wosiński, Krzysztof, *Bezpieczeństwo osób i systemów IT z wykorzystaniem białego wywiadu. OSINT*, PWN, Warszawa, 2024.
13. Wosiński, Krzysztof, *Google'a szkiełko i oko. Czyli co nowego w wyszukiwaniu zawartością obrazów [czwartki z OSINTem]*, https://sekurak.pl/googlea-szkielko-i-oko-czyli-co-nowego-w-wyszukiwaniu-zawartoscia-obrazow-czwartki-z-osintem/. Dostęp: 17.11.2024.
14. Wosiński, Krzysztof, *Wykorzystanie kont „alternatywnych tożsamości" w działaniach prowadzonych w Internecie*, Cybersecurity&Cybercrime, numer 2024/2 (4), Gdynia, 2024.