ZESZYTY NAUKOWE
WYŻSZEJ SZKOŁY BANKOWEJ W POZNANIU
2026, T. 110, NR 1

TOMASZ JANCZEWSKI
Uniwersytet WSB Merito w Poznaniu
Wyższa Szkoła Bankowa w Warszawie
https://orcid.org/0009-0006-4583-4377
tomasz@janczewski.it

Informatyka śledcza wspierana przez sztuczną inteligencję — od reaktywnego dochodzenia do predykcyjnej integralności dowodów

Streszczenie. Celem artykułu jest przedstawienie koncepcji zastosowania algorytmów sztucznej inteligencji w procesach informatyki śledczej, ze szczególnym uwzględnieniem wykrywania manipulacji danymi oraz zapewnienia integralności dowodów cyfrowych. Artykuł omawia koncepcję Zero Trust Digital Forensics i proponuje metodologię integracji zasad zero trust z procesami analizy dowodowej.

Słowa kluczowe: sztuczna inteligencja, informatyka śledcza, zero trust, integralność danych, dowody cyfrowe

https://doi.org/10.58683/dnswsb.2136

1. Wprowadzenie

Dynamiczny rozwój technologii cyfrowych i gwałtowny wzrost wykorzystania sztucznej inteligencji (AI) w środowiskach korporacyjnych i instytucjonalnych tworzą nową rzeczywistość dla praktyki informatyki śledczej. Tradycyjne modele analizy incydentów, oparte na zaufaniu do źródeł i manualnej weryfikacji danych, stają się niewystarczające w świecie, w którym przetwarzanie informacji w chmurze, automatyzacja decyzji oraz złożone środowiska pracy typu digital workspace tworzą nieprzewidywalne i podatne na manipulację ekosystemy cyfrowe. Jak zauważa Janczewski (2025c), współczesna informatyka śledcza wymaga przeformułowania swojej metodologii, ponieważ zaufanie do systemów, narzędzi i operatorów staje się coraz bardziej ryzykowne w kontekście bezpieczeństwa i integralności dowodów cyfrowych.

Według raportu NIST SP 800-207, tradycyjne modele bezpieczeństwa oparte na obwodach zaufania nie są w stanie sprostać współczesnym wyzwaniom związanym z rozproszoną infrastrukturą, złożonością środowisk chmurowych oraz mobilnością danych (Rose et al., 2020). Dlatego koncepcja Zero Trust Architecture (ZTA) — opierająca się na zasadzie „nigdy nie ufaj, zawsze weryfikuj" — stanowi fundament nowego paradygmatu w cyberbezpieczeństwie i coraz częściej w informatyce śledczej. Zgodnie z tą zasadą żaden element systemu nie może być uznany za wiarygodny bez weryfikacji jego tożsamości, integralności i kontekstu operacyjnego.

Z kolei raport *Gartnera Hype Cycle for Emerging Technologies 2025* wskazuje, że technologie autonomiczne i inteligentne systemy wspomagające decyzyjność (m.in. decision intelligence, AI agents, disinformation security) kształtują nową erę zarządzania informacją i ryzykiem w przedsiębiorstwach (Gartner, 2025). Gartner podkreśla, że wchodzimy w fazę technosocietal fragility — zjawiska, w którym wzrost uzależnienia od systemów cyfrowych czyni je jednocześnie niezbędnymi i wyjątkowo kruchymi. W tym kontekście automatyzacja procesów dochodzeniowych przy jednoczesnym utrzymaniu rygorystycznych zasad weryfikacji i integralności danych staje się kluczowa dla przyszłości informatyki śledczej.

Celem niniejszego artykułu jest przedstawienie potrzeby transformacji metod dochodzeniowych poprzez implementację zasad architektury Zero Trust i wykorzystanie sztucznej inteligencji w procesach analizy oraz walidacji dowodów cyfrowych. Artykuł koncentruje się na interdyscyplinarnym połączeniu informatyki śledczej, bezpieczeństwa IT i zarządzania danymi w środowiskach zautomatyzowanych, z naciskiem na rosnącą rolę AI w detekcji anomalii i potwierdzaniu autentyczności informacji. Zakres pracy obejmuje analizę trendów technologicznych (na podstawie raportów Gartnera), omówienie standardów NIST oraz zaprezentowanie autorskiego modelu weryfikacyjno-automatyzacyjnego inspirowanego zasadami Zero Trust Digital Forensics opracowanymi przez autora we wcześniejszej publikacji (Janczewski, 2025a).

Współczesne wyzwania w zakresie cyberbezpieczeństwa oraz zwiększona rola automatyzacji wymuszają przejście od modelu reaktywnego — polegającego na analizie zdarzeń po ich wystąpieniu — do modelu predykcyjnego, w którym systemy są zdolne do antycypowania zagrożeń oraz walidacji wiarygodności dowodów w czasie rzeczywistym. Tak zdefiniowany nowy paradygmat informatyki śledczej, łączący zasady Zero Trust i sztucznej inteligencji, stanowi istotny krok w kierunku budowy cyfrowego ekosystemu opartego na weryfikowalnym zaufaniu, automatyzacji i transparentności.

2. Przegląd literatury i podstawy teoretyczne

Badania nad zastosowaniem sztucznej inteligencji w analizie śledczej systemów informatycznych to dynamicznie rozwijający się obszar, który łączy aspekty informatyki śledczej, cyberbezpieczeństwa i automatyzacji procesów analitycznych. Współczesna informatyka śledcza nie może ograniczać się jedynie do rekonstrukcji zdarzeń — jej zadaniem staje się również przewidywanie i przeciwdziałanie incydentom bezpieczeństwa poprzez inteligentną analizę danych i ich kontekstu operacyjnego.

Jak zauważył Garfinkel (2010), przez długie lata podstawą informatyki śledczej była manualna analiza danych cyfrowych, której skuteczność ograniczała się do badania skutków, nie przyczyn zdarzeń. Dopiero rozwój uczenia maszynowego i metod opartych na sztucznej inteligencji umożliwił stworzenie modeli predykcyjnych pozwalających na identyfikację anomalii jeszcze przed wystąpieniem zdarzenia. W późniejszych latach badania, takie jak te prowadzone przez Neale'a i współautorów (2022), wprowadziły koncepcję Zero Trust Digital Forensics — modelu, w którym integralność dowodów cyfrowych nie jest zakładana z góry, lecz każdorazowo weryfikowana przez systemy oparte na politykach kontekstowych.

W klasycznym ujęciu informatyki śledczej kluczowym pojęciem jest *integralność dowodu cyfrowego*, rozumiana jako pewność, że dane nie uległy modyfikacji od momentu ich pozyskania. NIST w publikacji SP 800-207 wskazuje, że w architekturze Zero Trust każde źródło danych — w tym dane śledcze — musi przejść proces ciągłej autoryzacji i walidacji (Rose et al., 2020). Innymi słowy, integralność nie jest już stanem trwałym, lecz dynamicznym procesem potwierdzania wiarygodności danych.

Własne badania Janczewskiego (2025b) dotyczące automatycznej analizy logów serwerowych w kontekście wykrywania anomalii wykazały, że połączenie algorytmu Isolation Forest z regułami kontekstowymi umożliwia skuteczne wykrywanie nieprawidłowości w danych logów, przy jednoczesnym ograniczeniu liczby fałszywych alarmów. Wyniki te wpisują się w szerszy nurt badań, zgodnie z którym analiza logów przy użyciu sztucznej inteligencji nie tylko przyspiesza proces detekcji, ale też wspiera proces śledczy poprzez automatyczne wnioskowanie o potencjalnych przyczynach incydentu.

Z kolei Pissanidis i Demertzis (2023) analizują zastosowanie technologii Open XDR (Extended Detection and Response), wskazując, że wykorzystanie uczenia maszynowego w systemach korelacji zdarzeń pozwala łączyć dane z wielu źródeł w czasie rzeczywistym, co zwiększa zarówno skuteczność detekcji, jak i zdolność do automatycznej reakcji na zagrożenia. Ten trend jest zbieżny z kierunkiem rozwoju, jaki wyznaczają raporty Gartnera — w szczególności *Market Guide for Zero-Trust Network Access* (2025), w którym podkreślono, że technologie Zero Trust

nie mogą funkcjonować w oderwaniu od inteligentnych systemów analitycznych zapewniających ciągłe wnioskowanie o poziomie zaufania i ryzyku (Gao & Yu, 2025). Gartner wskazuje, że rozwiązania klasy ZTNA (Zero Trust Network Access) są obecnie integrowane z komponentami Security Service Edge (SSE), co umożliwia ich zastosowanie również w procesach dochodzeniowych i forensycznych.

Neale i współautorzy (2022) zauważają, że przejście od klasycznego modelu zaufania do modelu weryfikacyjnego wymaga nowego podejścia metodologicznego — takiego, które łączy automatyzację z nieprzerwanym nadzorem nad integralnością danych. W ich badaniach zaproponowano zintegrowane ramy analityczne, w których agent AI dokonuje wstępnej kwalifikacji danych dowodowych, a następnie system Zero Trust zatwierdza ich wiarygodność na podstawie kontekstu operacyjnego i metadanych.

W tym kontekście kluczowe staje się powiązanie zasad architektury Zero Trust z procesami analizy śledczej. Zgodnie z koncepcją przedstawioną przez NIST każdy komponent systemu informatycznego — w tym urządzenia, użytkownicy i dane — musi być traktowany jako potencjalnie niewiarygodny, dopóki jego autentyczność nie zostanie potwierdzona (Rose et al., 2020). Takie podejście zmienia rolę informatyki śledczej: z pasywnej analizy w kierunku aktywnego, zautomatyzowanego nadzoru nad integralnością dowodów.

Badania Garfinkela, Neale'a i Janczewskiego wskazują więc, że przyszłość informatyki śledczej leży w synergii sztucznej inteligencji, polityk bezpieczeństwa opartych na weryfikacji i architekturze Zero Trust. To połączenie pozwala nie tylko lepiej chronić dane, ale także skrócić czas reakcji śledczej i zwiększyć odporność środowisk cyfrowych na manipulacje.

3. Metodologia Zero Trust Digital Forensics (ZTDF)

Koncepcja Zero Trust Digital Forensics (ZTDF) stanowi rozwinięcie idei Zero Trust Architecture (ZTA) w kierunku metodologicznego ujęcia procesów analizy śledczej w środowiskach rozproszonych, opartych na automatyzacji i sztucznej inteligencji. Fundamentem tej metodologii jest założenie, że żaden komponent — system, użytkownik, aplikacja ani artefakt dowodowy — nie może być uznany za zaufany bez weryfikacji jego integralności, kontekstu oraz pochodzenia. ZTDF łączy zatem klasyczne zasady informatyki śledczej z mechanizmami dynamicznego uwierzytelniania, segmentacji procesów oraz ciągłej autoryzacji, tworząc model o charakterze warstwowym.

W odróżnieniu od tradycyjnych modeli dochodzeniowych, w których zakładano wiarygodność źródeł danych, metodologia ZTDF bazuje na ciągłym potwierdza-

niu tożsamości, integralności i autentyczności każdego komponentu środowiska badawczego. Takie podejście eliminuje element zaufania implicytnego — co jest zgodne z założeniami NIST SP 800-207, według którego „każdy podmiot w systemie musi być weryfikowany przed uzyskaniem dostępu do zasobów" (Rose et al., 2020).

4. Zasady metodologiczne

4.1. Weryfikacja wszystkich komponentów

W ramach ZTDF każdy element systemu — zarówno dane dowodowe, jak i narzędzia użyte w analizie — podlega wielostopniowej weryfikacji. Proces ten obejmuje m.in. haszowanie plików źródłowych, porównanie wyników z wzorcami kontrolnymi oraz rejestrację operacji w dzienniku integralności. Podobne podejście zaproponował Jilani (2025), integrując mechanizmy blockchain z procesem zabezpieczania dowodów, co umożliwiło pełną transparentność i odporność na manipulację.

W autorskiej implementacji ZTDF weryfikacja obejmuje nie tylko artefakty dowodowe, ale także modele uczenia maszynowego wykorzystywane w analizie. Modele AI, w tym algorytmy detekcji anomalii (np. Isolation Forest, LSTM), są testowane pod kątem stabilności wyników, by wykluczyć błędy spowodowane tzw. data drift — zjawiskiem zmiany statystycznych własności danych wejściowych.

4.2. Segmentacja procesów

Metodologia ZTDF zakłada rozdzielenie procesów analitycznych na izolowane segmenty funkcjonalne, co ogranicza propagację błędów i zapewnia pełną ścieżkę audytu. Inspiracją do tego podejścia są rozwiązania modułowe opisane w artykule *Enhancing Digital Forensics Investigations Using AI Driven Anomaly Detection and Log Correlation* (Future Engineering Journal, 2025), w których dane przechodzą przez sekwencję niezależnych etapów — od akwizycji, przez wstępne czyszczenie i korelację logów, po rekonstrukcję zdarzeń. Każdy segment analizy kończy się podpisaniem kryptograficznym i walidacją wyników poprzedniego etapu, co umożliwia pełne odtworzenie łańcucha dowodowego i minimalizuje ryzyko błędów interpretacyjnych (Felix, 2025).

Własne obserwacje autora, zawarte w publikacji *Implementacja praktyk DevSecOps w środowiskach chmurowych dla infrastruktury krytycznej* (2025a), wskazują, że segmentacja analityczna redukuje ryzyko eskalacji błędów oraz umożliwia skuteczniejsze wdrożenie zasad least privilege w kontekście operacji śledczych.

4.3. Dynamiczna autoryzacja

Jednym z kluczowych elementów ZTDF jest zastąpienie statycznych polityk dostępu mechanizmami dynamicznej autoryzacji opartej na kontekście. Każda operacja analityczna jest poprzedzana sprawdzeniem stanu komponentu (np. integralności systemu narzędziowego, wersji modelu AI, autentyczności użytkownika). Zgodnie z raportem Gartnera *Market Guide for Zero-Trust Network Access* (2025) to właśnie mechanizmy dynamicznego przydziału uprawnień stanowią trzon współczesnych architektur bezpieczeństwa, umożliwiając „ciągłe wnioskowanie o poziomie zaufania w czasie rzeczywistym" (Gao & Yu, 2025).

4.4. Model warstwowy ZTDF

Autorska metodologia ZTDF przyjmuje strukturę pięciu warstw funkcjonalnych, które odpowiadają za realizację zasad weryfikacji, segmentacji i autoryzacji:

- Warstwa identyfikacji — rozpoznaje i klasyfikuje artefakty dowodowe, użytkowników i procesy systemowe. W tym etapie stosuje się analizę metadanych i fingerprinting plików.
- Warstwa uwierzytelniania — potwierdza tożsamość komponentów poprzez wielopoziomowe mechanizmy kryptograficzne. Każdy element (dane, narzędzie, model AI) uzyskuje token dowodowy z unikalnym identyfikatorem czasowym.
- Warstwa kontroli dostępu — zarządza autoryzacją dynamiczną na podstawie kontekstu operacji, poziomu zaufania i aktualnej oceny ryzyka.
- Warstwa weryfikacji integralności — porównuje wartości hash, analizuje spójność danych z logami systemowymi oraz monitoruje zmiany w środowisku śledczym. Implementacja może wykorzystywać technologie blockchain, zgodnie z koncepcją ForensiBlock Akbarfama i współautorów (2023).
- Warstwa audytu i rekonstrukcji — zapewnia pełną ścieżkę inspekcji i możliwość odtworzenia procesu analizy. W tej fazie generowany jest raport śledczy z podpisem cyfrowym, a dane są walidowane przez niezależny mechanizm weryfikacji (np. serwer dowodowy w chmurze).

Schemat warstwowy ZTDF opiera się na założeniu, że każda operacja pozostawia ślad kryptograficzny, który umożliwia późniejsze odtworzenie sekwencji działań. W efekcie system pełni zarówno funkcję dochodzeniową, jak i kontrolną.

4.5. Integracja AI i automatyzacji w ZTDF

ZTDF implementuje mechanizmy uczenia maszynowego na trzech poziomach:

- detekcji anomalii (LSTM, Random Forest) — zgodnie z badaniami Felix (2025), które wykazały znaczący wzrost skuteczności w rekonstrukcji incydentów na podstawie skorelowanych logów;
- kontekstowej walidacji — wykorzystującej modele NLP do analizy opisów zdarzeń i ich semantyki;
- automatycznego triage'u dowodów — w oparciu o priorytetyzację ustaloną przez model decyzyjny AI.

Celem integracji AI w ZTDF nie jest zastąpienie analityka, lecz zautomatyzowanie procesów weryfikacyjnych i dostarczenie narzędzi predykcyjnych umożliwiających wcześniejsze wykrywanie nieprawidłowości. Jilani (2025) podkreśla, że kluczowym elementem skutecznej automatyzacji jest transparentność modeli (Explainable AI), bez której niemożliwe jest prawne uznanie wyników analizy za wiarygodne.

4.6. Weryfikacja integralności i ścieżka dowodowa

Integralność dowodu cyfrowego stanowi oś metodologiczną ZTDF. Każdy etap pracy analitycznej jest rejestrowany w sposób umożliwiający jego odtworzenie. W implementacji zaproponowanej przez autora dane są zapisywane w zdecentralizowanym rejestrze (blockchain), który pełni funkcję immutable audit trail. Podobne podejście zastosowała Bonomi i współautorzy (2018) w projekcie B-CoC (Blockchain-based Chain of Custody), umożliwiającym niezaprzeczalną rejestrację zdarzeń w procesach śledczych.

Własne badania autora nad wdrożeniem zasad Zero Trust w środowiskach informatyki śledczej dowiodły, że włączenie weryfikacji kryptograficznej oraz segmentacji logicznej procesów pozwala nie tylko zwiększyć odporność systemu na manipulację, ale również przyspieszyć procesy audytowe poprzez eliminację powtarzalnych czynności manualnych.

4.7. Znaczenie metodologii ZTDF

Metodologia ZTDF tworzy ramy, w których informatyka śledcza zyskuje nowy wymiar — od reaktywnej analizy do predykcyjnego nadzoru. Poprzez połączenie zasad Zero Trust, automatyzacji i analityki opartej na AI, możliwe staje się stworzenie środowiska, w którym każdy dowód jest weryfikowalny, a każdy etap anali-

zy — powtarzalny i transparentny. W ujęciu praktycznym ZTDF może stanowić fundament dla systemów automatycznej rekonstrukcji incydentów w środowiskach chmurowych oraz platform Security Operation Center (SOC) nowej generacji.

Jak zauważa Gartner (2025), przyszłość zarządzania bezpieczeństwem to „ciągłe dostosowywanie zaufania do dynamicznego kontekstu operacyjnego". W tym sensie ZTDF staje się naturalnym przedłużeniem idei Zero Trust w sferze informatyki śledczej — tworząc most pomiędzy automatyzacją a integralnością, pomiędzy sztuczną inteligencją a zaufaniem opartym na dowodach.

5. Zastosowanie algorytmów sztucznej inteligencji w analizie dowodów cyfrowych

W ramach metodologii Zero Trust Digital Forensics (ZTDF) kluczowym elementem pozostaje integracja algorytmów sztucznej inteligencji, które umożliwiają automatyzację procesów analitycznych i podniesienie poziomu wiarygodności dowodowej poprzez weryfikowalne modele uczenia maszynowego. AI w informatyce śledczej odgrywa dziś podwójną rolę: jest zarówno narzędziem analitycznym wspierającym identyfikację anomalii i manipulacji, jak i podmiotem wymagającym kontroli — ze względu na potrzebę zapewnienia przejrzystości decyzji.

5.1. Wykrywanie anomalii w logach systemowych

Analiza logów stanowi jedno z najważniejszych pól zastosowania AI w praktyce śledczej. W tradycyjnych metodach anomalię rozpoznaje analityk poprzez ręczną inspekcję zapisów systemowych. Jednak przy wolumenie danych generowanym przez współczesne środowiska chmurowe taka praktyka staje się niewykonalna. Modele uczenia maszynowego, takie jak Isolation Forest, LSTM (Long Short-Term Memory), oferują zdolność wykrywania subtelnych odchyleń w ciągach czasowych logów systemowych.

Jak wskazują Du i współautorzy (2020), w ramach badań nad zastosowaniem AI w informatyce śledczej LSTM umożliwia skuteczne rozpoznawanie wzorców sekwencyjnych w danych operacyjnych, co przekłada się na zdolność wykrywania nienaturalnych zmian w zachowaniu użytkowników i systemów. Własne badania autora (Janczewski, 2025b) wykazały, że zastosowanie modelu Isolation Forest w połączeniu z regułami kontekstowymi w analizie logów serwerowych pozwala zredukować liczbę fałszywych alarmów przy jednoczesnym zachowaniu wysokiej czułości detekcji.

W praktyce ZTDF wykrywanie anomalii stanowi pierwszy etap procesu walidacji — każda wykryta nieprawidłowość jest następnie kierowana do weryfikacji kontekstowej, a dopiero potem oznaczana jako potencjalny incydent. W ten sposób unika się sytuacji, w której decyzje modelu AI są przyjmowane bezkrytycznie, co jest zgodne z zasadą „nigdy nie ufaj, zawsze weryfikuj" opisanej w NIST SP 800-207.

5.2. Analiza metadanych i integralność dowodów

Sztuczna inteligencja odgrywa także coraz większą rolę w analizie metadanych, które stanowią istotny element procesu dowodowego. Metadane plików, zdjęć czy wiadomości mogą ujawniać informacje o czasie modyfikacji, pochodzeniu czy urządzeniu źródłowym. Jak wskazuje Billah (2025), zastosowanie technik Explainable AI (XAI) w informatyce śledczej umożliwia tworzenie interpretowalnych modeli, które wyjaśniają, dlaczego dana cecha (np. nietypowy znacznik czasu lub brak podpisu kryptograficznego) została uznana za istotną dla sprawy.

Systemy oparte na XAI, wykorzystujące metody SHAP i LIME, pozwalają analitykom nie tylko wykrywać odchylenia w metadanych, ale także zrozumieć, które atrybuty wpłynęły na decyzję klasyfikacyjną. To podejście jest szczególnie istotne w kontekście zgodności z wymogami sądowymi — tylko modele, których decyzje są w pełni wytłumaczalne, mogą zostać uznane za dopuszczalne w postępowaniu prawnym.

W koncepcji ZTDF każdy artefakt dowodowy przechodzi proces walidacji integralności, a metadane są analizowane przy użyciu autoenkoderów, które porównują ich wewnętrzną strukturę z modelem wzorcowym. Jak zauważyli Du i współautorzy (2020), takie rozwiązanie pozwala identyfikować anomalie w strukturze plików i wykrywać nieautoryzowane modyfikacje nawet w sytuacjach, gdy dane zostały częściowo nadpisane lub ukryte w pamięci masowej.

5.3. Wykrywanie manipulacji typu deepfake

Rozwój technik generatywnych, takich jak Generative Adversarial Networks (GANs), stworzył nowe wyzwania dla informatyki śledczej. Manipulacje typu deepfake mogą dotyczyć zarówno materiałów wideo, jak i dźwięku, co rodzi ryzyko wprowadzenia fałszywych dowodów. Badania Du i współautorów (2020) oraz innych naukowców wykazały, że głębokie sieci neuronowe mogą być z powodzeniem wykorzystywane zarówno do tworzenia, jak i wykrywania sfałszowanych treści.

Modele wykrywające deepfake — zwłaszcza te oparte na autoenkoderach konwolucyjnych — uczą się charakterystycznych zniekształceń wynikających z procesu generowania obrazu, takich jak artefakty na granicach twarzy czy nienaturalne

zmiany światła. W implementacji ZTDF tego rodzaju analiza stanowi część warstwy weryfikacji integralności, której zadaniem jest ocena autentyczności materiałów dowodowych z wykorzystaniem uczenia kontrastowego i porównań z bazą wzorców oryginalnych treści.

Własne obserwacje autora, oparte na analizie logów oraz danych wizualnych, potwierdzają, że zastosowanie autoenkoderów pozwala wykrywać subtelne manipulacje, które nie są zauważalne dla człowieka, a jednocześnie można je logicznie wytłumaczyć za pomocą metod XAI.

5.4. Automatyczna klasyfikacja dowodów

Klasyfikacja dowodów cyfrowych przy użyciu uczenia maszynowego staje się kluczowym elementem automatyzacji śledztw. W swoich badaniach nad analizą logów serwera Janczewski (2025b) wskazuje, że połączenie metod heurystycznych z algorytmami uczenia maszynowego może znacząco skrócić czas wstępnej analizy w porównaniu z podejściem manualnym.

W praktyce oznacza to, że systemy oparte na ZTDF potrafią automatycznie grupować artefakty w kategorie: dane sieciowe, zapisy logów, pliki multimedialne, dane użytkowników oraz dowody pośrednie. W modelu warstwowym ZTDF klasyfikacja taka jest częścią warstwy identyfikacji i ściśle współpracuje z modułami uwierzytelniania i audytu.

Badania Billaha (2025) potwierdzają, że integracja algorytmów CNN i LSTM z mechanizmami interpretacyjnymi SHAP i LIME umożliwia osiągnięcie wysokiej precyzji klasyfikacji (ponad 93%) przy jednoczesnym zachowaniu transparentności decyzji. Wyniki te są zgodne z raportem Gartnera (2025), który wskazuje, że modele hybrydowe łączące nadzorowane uczenie z przejrzystością stanowią przyszłość systemów detekcji i analizy cyfrowej.

5.5. Znaczenie integracji modeli AI w ramach ZTDF

W ujęciu holistycznym zastosowanie AI w ZTDF nie polega wyłącznie na detekcji anomalii, lecz na tworzeniu samokontrolującego się ekosystemu analitycznego. Każdy model (np. Isolation Forest, LSTM, autoenkoder) jest poddawany weryfikacji integralności w cyklu zamkniętym — analogicznie do weryfikacji komponentów w architekturze Zero Trust.

Jak podkreśla Billah (2025), przyszłość informatyki śledczej opiera się na łączeniu przejrzystości (XAI) i odporności operacyjnej, co stanowi warunek konieczny, by dowody oparte na analizie maszynowej mogły być uznane za prawnie wiarygodne. ZTDF w tym kontekście stanowi nie tylko metodologię analizy, lecz także

ramy etyczne — nakłada obowiązek pełnej przejrzystości procesu, kontroli danych wejściowych i zapewnienia powtarzalności wyników.

6. Dyskusja wyników i wnioski dla praktyki

Wprowadzenie metodologii Zero Trust Digital Forensics (ZTDF), opartej na zasadach weryfikacji, segmentacji i dynamicznej autoryzacji, wyznacza kierunek transformacji współczesnej informatyki śledczej. Dotychczasowe rozważania wykazały, że integracja sztucznej inteligencji (AI) z architekturą Zero Trust tworzy spójne ramy analityczne, które pozwalają zwiększyć efektywność pracy biegłych oraz wiarygodność materiału dowodowego. Jednak zastosowanie tych koncepcji w praktyce wymaga ostrożności i systemowego podejścia do kwestii standaryzacji, certyfikacji i nadzoru nad procesami automatyzacji.

6.1. Znaczenie koncepcji ZTDF dla praktyki dochodzeniowej

Zastosowanie zasad Zero Trust w informatyce śledczej oznacza odejście od tradycyjnego modelu, w którym dane pozyskane z urządzeń lub systemów były traktowane jako zaufane. ZTDF redefiniuje ten paradygmat: żaden artefakt, żadne narzędzie ani wynik działania algorytmu nie jest uznawany za wiarygodny bez uprzedniej weryfikacji integralności i źródła pochodzenia. Jak wskazuje Janczewski (2025c), praktyczne wdrożenie takiego podejścia wymaga projektowania śledczych procesów analitycznych w sposób modułowy — z rejestrowaniem każdego kroku w łańcuchu przetwarzania danych.

W praktyce dochodzeniowej ZTDF zmienia również rolę biegłego informatyka. Zamiast pełnić funkcję interpretatora danych, staje się on architektem systemu weryfikacyjnego, który nadzoruje automatyczne procesy walidacji i klasyfikacji dowodów. Każdy etap analizy — od akwizycji po raport końcowy — musi być nie tylko powtarzalny, ale także audytowalny. W tym kontekście model ZTDF stanowi próbę odpowiedzi na problem „czarnej skrzynki" algorytmów uczenia maszynowego, który ograniczał dotąd ich zastosowanie w postępowaniach sądowych (Janczewski, 2025c).

Jak wynika z badań Du i współautorów (2020), jednym z głównych wyzwań pozostaje brak pełnej przejrzystości modeli głębokiego uczenia w kontekście śledczym. Autorzy podkreślają, że dopiero połączenie AI z mechanizmami przejrzystości (XAI) oraz niezależnym systemem walidacji może zagwarantować, iż wynik analizy będzie zarówno skuteczny, jak i prawnie dopuszczalny.

6.2. Korzyści z wykorzystania sztucznej inteligencji

Zastosowanie AI w ramach metodologii ZTDF przynosi wymierne korzyści w trzech kluczowych obszarach: szybkości, precyzji i obiektywności analizy. Automatyczne wykrywanie anomalii w logach i metadanych umożliwia błyskawiczną identyfikację potencjalnych incydentów bez konieczności ręcznej selekcji danych. Modele takie jak Isolation Forest czy LSTM, odpowiednio trenowane i walidowane, pozwalają na wczesne wykrycie anomalii behawioralnych w infrastrukturze systemowej, co potwierdzają badania Janczewskiego (2025a) nad wykorzystaniem AI do analizy logów w środowiskach serwerowych.

Korzyścią dla praktyki jest również zdolność AI do korelowania danych z wielu źródeł — serwerów, urządzeń mobilnych i systemów chmurowych — co znacząco zwiększa wartość analityczną materiału dowodowego. Jak wskazuje Gartner (2025), systemy analityczne nowej generacji wykorzystujące uczenie głębokie osiągają obecnie poziom integracji danych niemożliwy do uzyskania w tradycyjnych narzędziach śledczych.

Istotnym aspektem jest też redukcja błędów ludzkich. Procesy automatyzacji pozwalają uniknąć błędnej interpretacji dowodów wynikającej z przeciążenia analityków lub złożoności materiału. Jak zauważył Billah (2025), modele wyjaśnialnej AI nie tylko dostarczają wyników, lecz także wskazują czynniki, które miały wpływ na klasyfikację — co umożliwia kontrolę procesu i weryfikację jego poprawności.

6.3. Ryzyka i ograniczenia automatyzacji

Choć automatyzacja procesów śledczych przyspiesza analizę, niesie również ryzyko systemowego błędu, który może zyskać skalę nieosiągalną w przypadku działań manualnych. Jak podkreśla Du i współautorzy (2020), nieprawidłowe dane wejściowe lub błędna interpretacja kontekstu przez model mogą prowadzić do powielania fałszywych wniosków w wielu przypadkach jednocześnie. Właśnie dlatego ZTDF zakłada, że każdy model AI musi być objęty nadzorem i okresową rewalidacją — analogicznie do testów narzędzi śledczych.

Ryzyko dotyczy również wiarygodności dowodowej. W postępowaniu sądowym wymaga się pełnej możliwości odtworzenia ścieżki analizy. Modele oparte na głębokich sieciach neuronowych (np. autoenkodery czy GAN) często generują wyniki trudne do jednoznacznej interpretacji. W konsekwencji nawet trafne ustalenia mogą być zakwestionowane, jeśli proces analizy nie jest transparentny.

Dodatkowym zagrożeniem jest nadmierne zaufanie do modeli decyzyjnych. Zgodnie z zasadą Zero Trust, automatyzacja nie może zastępować krytycznego

nadzoru człowieka. AI powinna wspierać proces dochodzeniowy, a nie go determinować. Praktyka pokazuje, że błędy interpretacyjne algorytmów zdarzają się szczególnie w przypadkach danych nieustrukturyzowanych, takich jak obrazy o niskiej jakości czy uszkodzone logi systemowe.

6.4. Konieczność standaryzacji i certyfikacji

Jednym z kluczowych wniosków wynikających z wdrażania ZTDF jest potrzeba stworzenia jednolitych standardów walidacji narzędzi i modeli AI stosowanych w analizie dowodowej. Jak wskazuje NIST w publikacji SP 800-207, każda implementacja systemu Zero Trust powinna być poprzedzona procesem oceny ryzyka i certyfikacji komponentów (Rose et al., 2020). W kontekście informatyki śledczej oznacza to konieczność opracowania norm technicznych i etycznych dotyczących wykorzystania uczenia maszynowego w procesach analizy i klasyfikacji.

Brak takich standardów powoduje rozbieżności w wynikach analiz i utrudnia ocenę wiarygodności dowodów w postępowaniu sądowym. Również Janczewski (2025a) podkreślał w swojej wcześniejszej pracy, że wdrożenie automatycznych systemów analitycznych w infrastrukturze krytycznej wymaga certyfikacji na poziomie zarówno oprogramowania, jak i procesów eksploatacyjnych.

Warto zauważyć, że problem standaryzacji ma również wymiar międzynarodowy. W krajach Unii Europejskiej trwają prace nad wprowadzeniem jednolitych zasad oceny narzędzi wykorzystujących sztuczną inteligencję w postępowaniach karnych. Projektowane regulacje, w tym propozycje dotyczące AI Act, zakładają konieczność okresowego audytu algorytmów oraz obowiązek przechowywania metadanych opisujących przebieg procesu decyzyjnego.

6.5. Wnioski dla praktyki

Analiza wyników badań i doświadczeń z implementacji ZTDF pozwala sformułować kilka praktycznych wniosków dla środowiska śledczego:

- AI jako narzędzie wspierające, nie decydujące — decyzje o charakterze procesowym powinny pozostać w gestii biegłego, natomiast algorytmy mogą pełnić funkcję doradczą.
- Transparentność i audytowalność — każdy etap działania modelu musi być dokumentowany w sposób umożliwiający jego odtworzenie.
- Rewalidacja modeli — podobnie jak narzędzia techniczne, modele uczenia maszynowego wymagają regularnych testów skuteczności i odporności na błędy.

- Współpraca interdyscyplinarna — rozwój metodologii ZTDF wymaga ścisłej współpracy informatyków śledczych, prawników i specjalistów od etyki danych.

Podsumowując, metodologia ZTDF stwarza fundament pod nowy model prowadzenia dochodzeń cyfrowych — taki, w którym zaufanie nie jest założeniem, lecz wynikiem weryfikacji. Integracja sztucznej inteligencji w procesach analizy dowodowej ma ogromny potencjał, ale wymaga rygorystycznych standardów jakości i certyfikacji, by mogła stać się trwałym elementem praktyki sądowej.

7. Podsumowanie i kierunki dalszych badań

Wprowadzenie koncepcji Zero Trust Digital Forensics (ZTDF) stanowi próbę odpowiedzi na jedno z najpoważniejszych wyzwań współczesnej informatyki śledczej: jak zapewnić wiarygodność i integralność dowodów w epoce zautomatyzowanych procesów, generatywnej sztucznej inteligencji i rozproszonych środowisk danych. Analiza przeprowadzona w poprzednich rozdziałach pokazuje, że połączenie zasad architektury Zero Trust z algorytmami uczenia maszynowego nie jest jedynie koncepcją teoretyczną, lecz realnym kierunkiem rozwoju metod śledczych, który umożliwia redefinicję pojęcia zaufania w procesach dochodzeniowych.

7.1. Znaczenie ZTDF dla rozwoju informatyki śledczej

Model ZTDF zakłada ciągłą weryfikację każdego elementu środowiska dowodowego — od źródeł danych po narzędzia analityczne i modele AI. Odejście od zaufania domyślnego i zastąpienie go weryfikacją integralności, autoryzacją kontekstową oraz pełnym audytem procesów wprowadza do praktyki śledczej nową jakość metodologiczną. W przeciwieństwie do tradycyjnych modeli, które koncentrowały się na rekonstrukcji zdarzeń po ich wystąpieniu, ZTDF umożliwia proaktywną analizę predykcyjną, wykrywającą anomalie i manipulacje w czasie rzeczywistym.

Jak wykazały badania Janczewskiego (2025a), automatyzacja procesów dochodzeniowych oparta na AI pozwala na istotne skrócenie czasu analizy i zwiększenie wykrywalności incydentów bez utraty wiarygodności wyników. To przesunięcie akcentu — z reaktywnego na predykcyjny charakter działań — wyznacza nowy etap rozwoju informatyki śledczej, w której kluczową rolę odgrywa analiza behawioralna systemów, a nie jedynie analiza statycznych danych.

Wyniki te są zgodne z obserwacjami Gartnera (2025), według którego przyszłość cyberbezpieczeństwa i analizy śledczej opiera się na „ciągłej rekontekstualizacji

zaufania" — dynamicznym przypisywaniu wiarygodności systemom i użytkownikom na podstawie aktualnych, a nie historycznych danych.

7.2. Wyzwania implementacyjne

Pomimo rosnącego potencjału ZTDF wdrożenie tego modelu w praktyce napotyka liczne bariery techniczne i organizacyjne. Po pierwsze, konieczne jest stworzenie zunifikowanych standardów dla modeli uczenia maszynowego wykorzystywanych w analizie dowodowej. Obecnie brak spójnych wytycznych dotyczących walidacji i certyfikacji algorytmów, co prowadzi do różnic w wynikach analiz nawet przy użyciu podobnych danych wejściowych.

Po drugie, jak podkreśla Rose i współautorzy w dokumencie NIST SP 800-207 (2020), skuteczność architektury Zero Trust zależy od stopnia dojrzałości procesów identyfikacji i weryfikacji w organizacji. W środowiskach śledczych oznacza to konieczność wdrożenia automatycznych mechanizmów uwierzytelniania, kryptograficznego podpisywania wyników oraz ciągłej walidacji integralności narzędzi.

Po trzecie, integracja AI w procesach dochodzeniowych rodzi pytania o odpowiedzialność i etykę. Kto ponosi odpowiedzialność za błąd algorytmu, który błędnie sklasyfikował dowód? Czy sąd może uznać wynik analizy opartej na modelu, którego działania nie można w pełni wyjaśnić? Jak wskazuje Billah (2025), tylko zastosowanie modeli Explainable AI (XAI) umożliwia zachowanie przejrzystości decyzji i ich zgodność z zasadami postępowania dowodowego.

7.3. Potencjał rozwoju i kierunki badań

W kontekście dalszych badań nad ZTDF można wyróżnić kilka kierunków, które wydają się kluczowe dla rozwoju tej dziedziny:

a. Automatyczna weryfikacja łańcucha dowodowego

Integracja technologii blockchain z procesami walidacji danych może zapewnić niezmienność i audytowalność łańcucha dowodowego. Badania Bonomi i współautorów (2018) potwierdzają, że zdecentralizowany zapis operacji analitycznych umożliwia budowę odpornego na manipulację rejestru dowodowego. W przyszłości ZTDF może wykorzystywać blockchain jako integralny komponent warstwy audytu i rekonstrukcji.

b. Modele hybrydowe AI

Połączenie klasycznych metod statystycznych z uczeniem głębokim (np. Isolation Forest + LSTM) pozwala uzyskać większą precyzję i stabilność analizy. Włas-

ne obserwacje autora sugerują, że modele hybrydowe charakteryzują się wyższą odpornością na błędy losowe i tzw. drift danych, co ma istotne znaczenie przy analizie długotrwałych zdarzeń sieciowych.

c. Weryfikowalne uczenie maszynowe (Verifiable ML)

Kolejnym kierunkiem rozwoju jest projektowanie modeli, które generują nie tylko wynik klasyfikacji, ale także matematyczne dowody poprawności swojego działania. Tego rodzaju koncepcje są rozwijane m.in. w środowiskach bezpieczeństwa krytycznego, gdzie każdy element procesu decyzyjnego musi być audytowalny i powtarzalny.

d. Standaryzacja procesów śledczych opartych na AI

Wdrożenie metodologii ZTDF na poziomie międzynarodowym wymaga stworzenia norm ISO lub EN definiujących minimalne wymagania dla narzędzi, modeli i procedur. Podobnie jak w przypadku klasycznych metod kryminalistyki, konieczna będzie akredytacja laboratoriów cyfrowych pod kątem zgodności z tymi standardami.

e. Integracja ZTDF z architekturą SOC i systemami SIEM

Włączenie komponentów ZTDF do platform zarządzania bezpieczeństwem (np. Splunk, Elastic Stack) pozwoliłoby połączyć bieżący monitoring z analizą śledczą. W rezultacie możliwe byłoby wykrywanie i śledzenie incydentów w czasie rzeczywistym, z pełnym zachowaniem łańcucha dowodowego.

7.4. Wnioski syntetyczne

Wdrożenie ZTDF oznacza przejście od klasycznego paradygmatu informatyki śledczej — skoncentrowanego na analizie post factum — do modelu dynamicznej weryfikacji opartego na zasadzie nieustannego potwierdzania zaufania. Integracja AI w tym procesie nie jest jedynie technologiczną innowacją, ale przede wszystkim narzędziem epistemologicznym, które zmienia sposób, w jaki rozumiemy dowód cyfrowy: z obiektu statycznego w dynamiczny proces logiczno-statystycznej walidacji.

Korzyści wynikające z tego podejścia — szybkość analizy, automatyzacja i zwiększona przejrzystość — są niezaprzeczalne. Jednocześnie jednak, jak pokazują doświadczenia praktyków, zbyt daleko idąca automatyzacja bez odpowiednich ram prawnych i etycznych może prowadzić do dehumanizacji procesu dowodowego. Dlatego kluczowe znaczenie ma równowaga pomiędzy automatyzacją a kontrolą ludzką oraz wprowadzenie mechanizmów certyfikacji modeli AI.

Ostatecznie ZTDF można uznać za pierwszy krok w kierunku budowy nowego paradygmatu informatyki śledczej — informatyki opartej na dowodzie weryfikowalnym, a nie domniemanym. Dalsze badania powinny koncentrować się na opracowaniu metod formalnego potwierdzania integralności modeli AI, rozwoju zdecentralizowanych rejestrów łańcucha dowodowego oraz budowie ekosystemów analitycznych zdolnych do autonomicznej, lecz audytowalnej analizy danych.

Bibliografia

Akbarfam, A.J., Heidaripour, M., Maleki, H., Dorai, G., & Agrawal, G. (2023). ForensiBlock: A Provenance-Driven Blockchain Framework for Data Forensics and Auditability. https://doi.org/10.48550/arXiv.2308.03927

Billah, M. (2025). Explainable AI for Digital Forensics: Ensuring Transparency in Legal Evidence Analysis. *Journal of Forensic Science Research, 9*(2), 109–116.

Bonomi, S., Casini, M., & Ciccotelli, C. (2018). B-CoC: A Blockchain-based Chain of Custody for Evidence Management in Digital Forensics. https://doi.org/10.4230/OASIcs.Tokenomics.2019.12

Du, X., Hargreaves, C., Sheppard, J., Anda, F., Sayakkara, A., Le-Khac, N.-A., & Scanlon, M. (2020). *SoK: Exploring the State of the Art and the Future Potential of Artificial Intelligence in Digital Forensic Investigation*. ARES Conference Proceedings, ACM.

Felix, A.O. (2025). Enhancing Digital Forensics Investigations Using AI Driven Anomaly Detection and Log Correlation: A Mixed Methods Approach. *International Journal of Future Engineering Innovations*, *2*(4), 71–84.

Gao, F., & Yu, M. (2025). Market Guide for Zero-Trust Network Access (ZTNA). *Gartner Research*, G00838074.

Garfinkel, S.L. (2010). Digital Forensics Research: The Next 10 Years. *Digital Investigation*, 7, S64–S73.

Gartner. (2025). *Hype Cycle for Emerging Technologies 2025*, G00824262.

Janczewski, T. (2025a). Implementacja praktyk DevSecOps w środowiskach chmurowych dla infrastruktury krytycznej. Analiza bezpieczeństwa procesów CI/CD. *Zeszyty Naukowe Wyższej Szkoły Bankowej w Poznaniu, 108*(1). https://doi.org/10.58683/dnswsb.2064

Janczewski, T. (2025b). Wykorzystanie sztucznej inteligencji do analizy logów serwera w celu wykrywania anomalii. *Przestępczość Teleinformatyczna, 1*.

Janczewski, T. (2025c). *Zastosowanie architektury Zero Trust w informatyce śledczej: nowe podejście metodologiczne do zapewnienia integralności dowodów cyfrowych*. Akademia Marynarki Wojennej w Gdyni.

Jilani, G.S. (2025). AI-Driven Cyber Forensics: Enhancing Digital Investigation and Threat Detection. *JETIR*, *12*(7).

Neale, C., Kennedy, I., Price, B., Yu, Y., & Nuseibeh, B. (2022). The Case for Zero Trust Digital Forensics. *Forensic Science International: Digital Investigation*, *40*, 301352.

Pissanidis, D., & Demertzis, K. (2023). *Integrating AI/ML in Cybersecurity: An Analysis of Open XDR Technology and Its Application in Intrusion Detection and System Log Management*. Preprints.org.

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). *Zero Trust Architecture*. NIST Special Publication 800-207. National Institute of Standards and Technology.

Ai-Driven Digital Forensics: From Reactive Investigation to Predictive Evidence Integrity

Abstract. The article presents the concept of applying artificial intelligence algorithms in digital forensics processes, focusing on detecting data manipulation and ensuring the integrity of digital evidence. It discusses the emerging concept of "Zero Trust Digital Forensics" and proposes a methodological framework for integrating zero trust principles into forensic investigation workflows.

Keywords: artificial intelligence, digital forensics, zero trust, data integrity, digital evidence