

DetectAI – instrukcja działania i struktura systemu 3.0

=====

Wersja 3.0 – rozszerzona, aktualizacja 2026-09-08

1. CZYM JEST DETECTAI

DetectAI jest silnikiem analizy dowodów (AI Evidence Engine), a nie pojedynczym klasyfikatorem „AI / człowiek”.

System zbiera niezależne ślady pochodzące z różnych źródeł:

- HTML strony,
- treści tekstowej,
- deklaracji o użyciu AI,
- nazw providerów i modeli,
- struktury dokumentu,
- metadanych plików,
- EXIF / XMP / IPTC,
- C2PA / Content Credentials,
- SynthID,
- innych technicznych oznaczeń pochodzenia.

Każdy moduł zwraca obiekty Evidence.

Typy wyników:

- Confirmed
- StrongEvidence
- Evidence
- NoEvidence
- Inconclusive

Najważniejsza zasada DetectAI:
BRAK DOWODU NIE JEST DOWODEM BRAKU AI.

2. GŁÓWNY PIPELINE STRONY WWW

Adres URL

- > HttpPageFetcher
- > SourceDetector
- > TextDetector
- > ProviderDetector
- > AnalysisWorkspace
- > ImageDetector
 - > MetadataDetector
 - > C2PADetector
 - > SynthIDDetector
- > EvidenceCollection
- > Raport DetectAI

Strona jest pobierana tylko raz. Ten sam PageFetchResult jest przekazywany do detektorów HTML, dzięki czemu wszystkie moduły analizują tę samą wersję strony.

3. TEXTDETECTOR – PODEJŚCIE 3.0

TextDetector nie odpowiada na pytanie „czy AI napisało ten tekst na pewno?”. Bada, czy tekst posiada zestaw cech stylometrycznych, strukturalnych i językowych, które często współwystępują w treściach generowanych lub silnie redagowanych przez modele językowe.

TextDetector jest warstwą heurystyczną, a nie technicznym dowodem pochodzenia.

4. PRZYGOTOWANIE TEKSTU

Dla HTML:

- wydobywana jest główna treść,
- ograniczany jest boilerplate,
- usuwane są duplikaty bloków,
- normalizowany jest Unicode,
- usuwane są znaki zero-width,
- porządkowany jest whitespace,
- zachowywane są logiczne bloki i akapity.

Dla plików lokalnych szczególnie ważne jest zachowanie struktury akapitów. W DOC/DOCX tekst nie powinien być składany do jednego wielkiego bloku.

5. CO ANALIZUJE TEXTDETECTOR

TextDetector bada m.in.:

- liczbę słów i zdań,
- średnią długość zdania,
- zmienność długości zdań,
- długość i regularność akapitów,
- powtarzalne początki zdań,
- frazy przejściowe,
- charakterystyczne frazy,
- MATTR,
- powtarzalność n-gramów,
- strukturę nagłówków,
- relację nagłówek -> opis,
- udział krótkich bloków,
- CTA,
- frazy marketingowe,
- formalne i proceduralne konstrukcje językowe.

StrongEvidence wymaga współwystępowania kilku niezależnych cech.

6. PROFILE TEXTDETECTORA

Główne profile:

- Long-form,

- Marketing / Landing Page,
- Formal Document,
- Mixed.

7. LONG-FORM

Profil dla artykułów, blogów, analiz i dłuższych treści informacyjnych. Analizowane są m.in. regularność zdań, zmienność, powtarzalność, frazy przejściowe, MATTR i n-gramy.

8. MARKETING / LANDING PAGE

Profil marketingowy bada:

- liczbę nagłówków,
- frazy marketingowe,
- układ nagłówek -> opis,
- CTA,
- udział krótkich bloków,
- semantyczne echo,
- powtarzanie podobnych obietnic innymi słowami.

Typowy styl marketingowy nie jest sam w sobie StrongEvidence.

9. NOWY PROFIL FORMAL DOCUMENT

To jedna z najważniejszych zmian TextDetector 2.0.

Profil został dodany dla:

- dokumentów instytucjonalnych,
- regulaminów,
- procedur,
- opisów działalności,
- dokumentów administracyjnych,
- treści silnie modularnych i formalnych.

Wcześniej taki tekst mógł być niedoszacowany, bo nie jest marketingowy, a jednocześnie często ma krótkie akapity, wiele nagłówków i bardzo regularną strukturę.

10. SYGNAŁY FORMAL DOCUMENT

Analizowane są m.in.:

- liczba formalnych fraz,
- liczba różnych formalnych fraz,
- zagęszczenie fraz na 1000 słów,
- regularność sekcji,
- relacja nagłówek -> treść,
- mikro-modularne akapity,
- korelacja formalnego języka ze strukturą.

Przykładowe konstrukcje:

- „zapewnienie odpowiednich...”
- „w ramach działalności...”
- „ma na celu...”
- „kierowane są do...”
- „działania na rzecz...”
- „stanowi element...”
- „kluczowe znaczenie ma...”
- „odpowiedzialność za...”

Sama pojedyncza fraza nie wystarcza do StrongEvidence.

11. DLACZEGO PROFIL FORMALNY JEST LEPSZY

Najważniejszą zmianą jest korelacja cech.

Silniejszym sygnałem jest jednoczesne wystąpienie:
języka formalnego + regularnej struktury + krótkich akapitów
niż każda z tych cech osobno.

To ogranicza przypadki, w których system reaguje zbyt mocno na jeden prosty wzorzec.

12. OGRANICZENIA PROFILU FORMALNEGO

Dokument napisany przez człowieka może również być bardzo regularny i formalny. Dlatego styl formalny nie jest dowodem AI.

Jeśli benchmark pokaże zbyt wiele false positive, właściwym kierunkiem jest:

- podniesienie progu StrongEvidence,
- zwiększenie liczby wymaganych sygnałów,
- wymaganie dodatkowego potwierdzenia stylometrycznego.

13. CONFIDENCE

Confidence nie oznacza „procentowego prawdopodobieństwa autorstwa AI”. Oznacza siłę i spójność wykrytych sygnałów heurystycznych.

14. ANALIZA PLIKÓW LOKALNYCH – NOWA FUNKCJA

DetectAI nie ogranicza się już wyłącznie do adresów URL.

Plik może zostać:

- wybrany przyciskiem Analize File,
- przeciągnięty do okna programu metodą drag & drop.

Po wskazaniu pliku program dobiera właściwą ścieżkę analizy.

15. FORMATY TEKSTOWE

TXT / MD / CSV / JSON / XML

- > odczyt tekstu
- > TextDetector
- > EvidenceCollection
- > raport

16. HTML / HTM

Lokalny HTML:

- > odczyt HTML
- > analiza głównej treści
- > TextDetector.AnalyzeHtmlAsync
- > raport

17. DOC / DOCX

DOC/DOCX:

- > Microsoft Word COM
- > iteracja po Paragraphs
- > zachowanie granic akapitów
- > TextDetector
- > raport

Zachowanie akapitów poprawia analizę:

- długości bloków,
- mikro-modularności,
- struktury formalnej,
- relacji sekcja -> treść.

18. PDF

PDF:

- > ekstrakcja tekstu
- > TextDetector
- > raport

W obecnym podejściu ekstrakcja może korzystać z Microsoft Word.

Ograniczenie:

zeskanowany PDF bez warstwy tekstowej nie jest automatycznie OCR-owany.

19. DRAG & DROP

Po upuszczeniu pliku:

1. program sprawdza rozszerzenie,
2. wybiera odpowiedni sposób ekstrakcji,
3. uruchamia TextDetector,
4. tworzy Evidence,
5. generuje raport.

20. CO POPRAWIŁO SIĘ W TEXTDETECTORZE

Najważniejsze ulepszenia:

- zachowanie akapitów,
- profilowanie tekstu,
- osobny profil Marketing,
- osobny profil Formal Document,
- rozszerzony katalog fraz formalnych,
- scoring zależny od zagęszczenia fraz,
- analiza liczby różnych formalnych konstrukcji,
- mikro-modularność,
- poprawiona relacja nagłówków -> opis,
- korelacja kilku sygnałów,
- zachowanie wysokiego progu StrongEvidence.

21. NA ILE TEXTDETECTOR JEST LEPSZY

Nie można uczciwie podać procentowej poprawy bez benchmarku.

Największa poprawa dotyczy:

- dokumentów formalnych,
- tekstów modularnych,
- DOC/DOCX,
- landing pages,
- analizy struktury akapitów.

Nadal trudne są:

- krótkie teksty,
- AI po ręcznej redakcji,
- bardzo szablonowe teksty ludzkie,
- dokumenty prawne,
- teksty mieszane człowiek + AI.

22. SOURCEDETECTOR

SourceDetector wyszukuje jawne deklaracje użycia AI w:

- treści,
- metatagach,
- JSON-LD,
- komentarzach,
- atrybutach,
- skryptach inline.

Istotna poprawa: odróżnianie zwykłej wzmianki o AI od deklaracji, że analizowany materiał został utworzony przez AI.

23. PROVIDERDETECTOR

ProviderDetector identyfikuje technologie i dostawców AI.

Sama nazwa providera jest słabszym sygnałem niż jawna deklaracja użycia.

Jeżeli strona wymienia wielu providerów, może być zwykłym materiałem redakcyjnym
o AI, więc waga takich trafień jest obniżana.

24. IMAGEDETECTOR

ImageDetector:

- znajduje obrazy,
- pobiera je bezpiecznie,
- kontroluje redirecty,
- chroni przed SSRF i DNS rebinding,
- sprawdza magic bytes,
- liczy SHA-256,
- deduplikuje,
- przekazuje ten sam plik do MetadataDetector, C2PADetector i SynthIDDetector.

25. METADATADETECTOR

Analizuje EXIF, XMP, IPTC, Software, Creator i inne pola.
Brak metadanych nie oznacza braku AI.

26. C2PA / CONTENT CREDENTIALS

C2PADetector analizuje nie tylko obecność manifestu, ale też:

- claimy,
- digitalSourceType,
- błędy walidacji,
- szczegółowe dane provenance.

Sam manifest nie jest automatycznie dowodem AI.
Krytyczne błędy walidacji obniżają siłę wyniku.

27. SYNTHID

SynthIDDetector nie zgaduje watermarka.
Do potwierdzenia potrzebny jest zaufany verifier.

Bez verifera wynik pozostaje Inconclusive, a nie NoEvidence.

28. AI W INNYCH FORMACH

Zasada rozwoju:

dowód techniczny -> metadata -> provenance -> watermark -> korelacja ->
heurystyka

Obraz:

priorytetem są C2PA, metadata, watermark i ślady providera.

Audio:

docelowo metadata, encoder, provenance, C2PA i watermark verifier.

Wideo:

metadata, C2PA, encoder, provenance, watermarki, analiza klatek i ścieżki audio.

29. DLACZEGO ALGORYTMY SĄ LEPSZE

Najważniejsza poprawa polega na ograniczaniu sytuacji, w których system wyciąga zbyt silny wniosek ze zbyt słabego sygnału.

Przykłady:

- SourceDetector nie myli polityki AI z deklaracją autorstwa.
- ProviderDetector nie myli artykułu o AI z użyciem wszystkich providerów.
- TextDetector korzysta z profili zamiast jednego scoringu.
- Formal Document wymaga korelacji kilku cech.
- DOC/DOCX zachowuje akapity.
- ImageDetector przekazuje te same bajty do całego pipeline.
- C2PA interpretuje znaczenie manifestu.
- SynthID wymaga verifera.
- Inconclusive nie jest zamieniane na NoEvidence.

30. WALIDACJA SKUTECZNOŚCI

Aby podawać procentową skuteczność, potrzebny jest benchmark zawierający:

- teksty ludzkie,
- teksty AI,
- AI po redakcji,
- teksty ludzkie poprawiane przez AI,
- marketing,
- dokumenty formalne,
- artykuły,
- różne długości i różne modele.

Należy mierzyć:

- true positive,
- false positive,
- true negative,
- false negative,
- precision,
- recall,
- specificity.

31. PODSUMOWANIE

DetectAI 3.0 rozwija projekt w dwóch głównych kierunkach.

Lepsza analiza treści:

- nowy TextDetector,

- profile tekstowe,
- Formal Document,
- mikro-modularność,
- zachowanie akapitów,
- lokalne DOC/DOCX/PDF/HTML/TXT,
- drag & drop,
- Analize File.

Mocniejsze dowody techniczne:

- bezpieczniejszy pipeline obrazów,
- SHA-256,
- rozwinięte C2PA,
- ostrożniejszy ProviderDetector,
- brak zgadywania SynthID,
- zachowanie Inconclusive.

DetectAI ma odpowiadać:

- jakie ślady znaleziono,
- skąd pochodzą,
- jak są silne,
- czy potwierdzają się nawzajem,

a nie jedynie zwracać jeden procent „AI”.

32. ZMIANY TEXTDETECTORA W WERSJI 3.0

Wersja 3.0 rozwija analizę tekstu z pojedynczych heurystyk do korelacji kilku niezależnych rodzin cech.

Najważniejsze nowe lub poprawione elementy:

- lepsze rozdzielenie akapitów od elementów list,
- ostrożniejsze rozpoznawanie nagłówków w plain text,
- bezpieczniejsze filtrowanie diagramów ASCII / box-drawing,
- osobna analiza szablonów ofert technicznych,
- równoległość gramatyczna list,
- rozszerzony Syntactic Flatness,
- Transition Frame Signal,
- struktura intro -> lista -> domknięcie,
- cytowania numeryczne RAG-like,
- kontrola obecności bibliografii / sekcji źródeł,
- symetria modułów nagłówek -> opis,
- echo leksykalne pomiędzy sekcjami,
- korelacja wielu rodzin cech,
- dynamiczny Confidence zależny od wyniku i ilości materiału.

33. AKAPITY I ELEMENTY LIST

Element listy nie jest już automatycznie liczony jako akapit.

Statystyki akapitów są liczone głównie dla:

- Paragraph,

- Quote.

Elementy ListItem nadal uczestniczą w analizie:

- struktury dokumentu,
- modularności,
- profilu formalnego,
- równoległości list,
- framingu list.

Zmiana zapobiega sytuacji, w której dokument z kilkudziesięcioma punktami raportuje dziesiątki bardzo krótkich „akapitów” i sztucznie zwiększa mikro-modularność.

34. OSTROŻNIEJSZE ROZPOZNAWANIE NAGŁÓWKÓW

W plain text nadal rozpoznawane są:

- nagłówki numerowane,
- krótkie nagłówki pisane wielkimi literami,
- samodzielne nagłówki typu:
Podsumowanie, Wnioski, Zakres, Wstęp, Summary, Conclusion.

Ogólny heurystyczny nagłówek został zaostrożony.

Celem jest ograniczenie sytuacji, w których:

- nazwy urzędzeń,
- etykiety techniczne,
- krótkie pozycje tabel,
- fragmenty parametrów

są błędnie klasyfikowane jako Heading.

35. DIAGRAMY ASCII / BOX-DRAWING

TextNormalization usuwa z analizy stylometrycznej linie wyglądające jak:

- diagramy ASCII,
- box-drawing,
- proste schematy tekstowe,
- wybrane tabele tekstowe.

Filtrowanie jest obecnie bardziej konserwatywne.

Program nie usuwa całej linii tylko dlatego, że znajduje się w niej kilka znaków schematu. Jest to ważne dla tekstu wklejanego, gdzie zwykłe zdanie może zostać spłaszczone do tej samej linii co fragment diagramu.

Krótkie etykiety są usuwane tylko wtedy, gdy wyraźnie znajdują się wewnątrz struktury diagramu.

36. SZABLONY OFERT TECHNICZNYCH

TextDetector posiada osobny katalog konstrukcji spotykanych w ofertach technicznych i opisach architektury.

Przykłady:

- „w związku z”,
- „ze względu na”,
- „proponujemy zastosowanie”,
- „proponujemy zachowanie”,
- „nie rekomendujemy migracji”,
- „w przedstawionej architekturze”,
- „rekomendujemy pozostawienie”,
- „proponowana architektura zapewnia”,
- „pozwala zachować wszystkie funkcje”,
- „wykorzystanie istniejącej infrastruktury”,
- „system pozostaje nadrzędnym”,
- „system odpowiada za”.

WAŻNE:

te konstrukcje NIE są sygnaturą AI.

Są cechą profilu formalno-technicznego i mają znaczenie dopiero w korelacji z innymi właściwościami dokumentu.

37. SYNTACTIC FLATNESS 3.0

Syntactic Flatness mierzy regularność długości zdań w sekcjach tekstu.

Wersja 3.0:

- ocenia sekcje od co najmniej 3 zdań,
- analizuje przesuwne okna 4 kolejnych zdań,
- mierzy współczynnik zmienności długości zdań,
- kontroluje rozpiętość długości zdań w oknie.

Jeżeli sekcje są bardzo krótkie, wykonywany jest dodatkowy pomiar na całym strumieniu prozy.

Dzięki temu układ:

nagłówek -> 1-2 zdania -> kolejny nagłówek

nie powoduje automatycznie wyniku „0 okien” tylko dlatego, że pojedyncze sekcje są krótkie.

Niska zmienność zdań nie jest samodzielnym dowodem AI.

38. RÓWNOLEGŁOŚĆ GRAMATYCZNA LIST

TextDetector analizuje grupy kolejnych elementów list.

Przykładowy wzorzec:

- obsługę...
- zarządzanie...
- kontrolę...
- definiowanie...

Detektor mierzy dominujący sposób gramatycznego otwarcia punktów.

Jest to heurystyka morfologiczna, a nie pełny tagger części mowy.

Raport może pokazywać:

- liczbę elementów list,
- liczbę grup z równoległym otwarciem,
- najwyższy poziom zgodności otwarć.

Silna symetria list jest sygnałem szablonowości, ale sama nie oznacza AI.

39. TRANSITION FRAME SIGNAL

Wersja 3.0 posiada osobny katalog szablonowych zapowiedzi list.

Przykłady:

- „oto kluczowe powody”,
- „oto najważniejsze”,
- „z następujących powodów”,
- „poniżej przedstawiono”,
- „należy zwrócić uwagę na następujące”.

Detektor nie ogranicza się do sprawdzenia obecności frazy.

Sprawdza również, czy po zapowiedzi rzeczywiście występuje:

- lista punktowana,
- lista numerowana,
- albo seria numerowanych modułów nagłówek -> opis.

40. STRUKTURA INTRO -> LISTA -> DOMKNIĘCIE

Nowa metryka rozpoznaje pełniejszy wzorzec odpowiedzi:

wprowadzenie

- > zapowiedź
- > lista
- > dalszy akapit / podsumowanie / wnioski

Raport rozróżnia:

- liczbę szablonowych zapowiedzi list,
- liczbę sekwencji zapowiedź -> lista,
- liczbę sekwencji zapowiedź -> lista -> domknięcie.

Jest to silniejsza cecha strukturalna niż sama obecność jednej frazy.

41. CYTOWANIA NUMERYCZNE / RAG-LIKE

TextDetector rozpoznaje wzorce:

- [1]
- [1, 2]
- [1, 2, 3]
- [1; 2; 3]

Jednocześnie sprawdza, czy tekst posiada sekcję typu:

- Bibliografia,
- Źródła,
- References,
- Sources.

Raport może pokazywać:

- liczbę grup cytowań,
- liczbę cytowań wieloźródłowych,
- liczbę cytowań bez wykrytej bibliografii,
- informację o obecności sekcji źródeł.

Znacznik [1, 2, 3] nie jest technicznym dowodem użycia RAG lub LLM. Jest wyłącznie sygnałem formatu, który nabiera znaczenia w korelacji z innymi cechami.

42. SYMETRIA NAGŁÓWEK -> OPIS

TextDetector mierzy, czy kolejne moduły mają podobną długość.

Przykład:

1. Nagłówek
opis
2. Nagłówek
opis
3. Nagłówek
opis

Analizowane są:

- liczba kwalifikujących się modułów,
- liczba modułów o podobnej długości,
- udział modułów podobnych,
- współczynnik zmienności długości opisów.

Do porównania używana jest również mediana długości opisów, aby ograniczyć wpływ pojedynczych skrajnych bloków.

43. ECHO LEKSYKALNE MIĘDZY SEKCJAMI

Wersja 3.0 analizuje podobieństwo słownictwa pomiędzy niezależnymi sekcjami.

Jest to analiza leksykalna, a nie model embeddingowy.

Detektor:

- dzieli tekst na sekcje,
- wybiera istotniejsze tokeny,
- porównuje zbiory słów,
- mierzy podobieństwo między parami sekcji.

Raport pokazuje:

- liczbę par sekcji z podwyższonym podobieństwem,
- maksymalne wykryte podobieństwo.

Cecha może wskazywać na:

- powtarzanie tych samych idei,
- encyklopedyczne echo,
- silną szablonowość.

Nie jest samodzielnym dowodem AI.

44. KORELACJA WIELU RODZIN CECH

Najważniejszym kierunkiem wersji 3.0 jest korelacja.

Silniejszy wynik powinien powstawać dopiero wtedy, gdy jednocześnie występują różne rodziny cech, np.:

- regularny rytm,
- framing list,
- cytowania RAG-like,
- symetria nagłówków -> opis,
- echo leksykalne,
- formalny język,
- równoległe listy.

Detektor nie powinien klasyfikować tekstu wysoko tylko dlatego, że znalazł jedną frazę lub jeden format cytowania.

45. CONFIDENCE 3.0

Confidence nadal NIE oznacza:

„procentowego prawdopodobieństwa, że tekst napisało AI”.

Confidence oznacza stabilność klasyfikacji heurystycznej.

W starszej formule wiele różnych dokumentów mogło otrzymywać podobne wartości, np. 42%, 46% albo 64%, ponieważ wynik opierał się głównie

na kilku progach binarnych.

W wersji 3.0 Confidence zależy płynnie od:

- końcowego score profilu,
- odległości score od progu klasyfikacji,
- liczby aktywnych sygnałów,
- liczby słów,
- liczby zdań,
- liczby bloków treści.

Dzięki temu różne teksty nie powinny otrzymywać tej samej wartości tylko dlatego, że przekroczyły te same progi.

46. CONFIDENCE DLA NOEVIDENCE

Dla NoEvidence logika jest odwrotna niż dla Evidence.

Im dalej wynik znajduje się poniżej progu Evidence, tym stabilniejsza może być klasyfikacja NoEvidence.

Nie oznacza to jednak:
„tekst na pewno napisał człowiek”.

NoEvidence oznacza wyłącznie:
bieżący moduł nie znalazł wystarczającego zestawu cech.

47. CONFIDENCE WPISÓW POMOCNICZYCH

Wpisy pomocnicze Evidence nie używają już prostego schematu:

45 + liczba punktów sygnału.

Wcześniej prowadziło to do praktycznie stałych wartości, np.:

- +13 pkt -> 58%,
- +12 pkt -> 57%.

W wersji 3.0 Confidence pojedynczego wpisu zależy od:

- Confidence całej klasyfikacji,
- końcowego score,
- siły konkretnego sygnału.

Ten sam sygnał może więc mieć inną wartość w słabszym i mocniejszym dokumencie.

48. ANALIZA WKLEJONEGO TEKSTU

Ścieżka analizy:

wklejony tekst

- > przygotowanie struktury
- > TextDetector.AnalyzeText
- > EvidenceCollection
- > raport

Warstwa formularza nie powinna sztucznie ustawiać:

- Evidence,
- StrongEvidence,
- Confidence,
- końcowego score.

Wynik pochodzi z TextDetector.

Dla tekstu wklejanego szczególnie ważne są:

- zachowanie list,
- odbudowa nagłówków,
- rozdzielenie akapitów,
- ostrożne filtrowanie diagramów,
- syntactic flatness,
- framing list,
- profil formalno-techniczny.

49. PEŁNY PIPELINE C2PA DLA URL

Przy analizie adresu URL C2PA jest uruchamiane wewnątrz ImageDetector.

Schemat:

URL

- > ImageDetector
 - > pobranie obrazu
- > MetadataDetector
- > C2PADetector
- > SynthIDDetector

C2PADetector pracuje na lokalnym pliku obrazu pobranym przez ImageDetector.

Wynik C2PA jest następnie dołączany do wspólnej EvidenceCollection.

50. C2PATool – OCZEKIWANA LOKALIZACJA

Na Windows C2PADetector domyślnie oczekuje:

<AppContext.BaseDirectory>\\Tools\\c2patool.exe

Typowa struktura po publikacji:

DetectAI.exe
Tools\\
c2patool.exe

Samo istnienie c2patool.exe w tym katalogu nie oznacza, że narzędzie zostało uruchomione.

C2PADetector musi zostać wywołany przez pipeline obrazu.

51. OGRANICZENIE LOKALNYCH OBRAZÓW

Obecna ścieżka Analyze File jest przede wszystkim ścieżką tekstową.

Okno pozwala wskazać także pliki:

- JPG,
- JPEG,
- PNG,
- WebP,
- GIF,
- BMP,
- TIFF.

Jednak pojedynczy lokalny obraz wskazany przez Analyze File nie przechodzi obecnie pełnego pipeline:

MetadataDetector -> C2PADetector -> SynthIDDetector.

Pełny pipeline obrazowy działa podczas analizy strony WWW.

Jest to znane ograniczenie wersji 3.0.

52. ORGANIZACJA MODUŁÓW TEKSTOWYCH

TextDetector.cs

- główna analiza,
- scoring,
- klasyfikacja,
- Confidence,
- opis metryk i Evidence.

TextNormalization.cs

- Unicode,
- whitespace,
- filtrowanie diagramów ASCII / box-drawing.

TextPhraseCatalog.cs

- frazy przejściowe,
- AI-tells,
- marketing,
- język formalny,
- szablony ofert technicznych,
- zapowiedzi list.

TextRegexCatalog.cs

- wspólne regexy,
- listy,
- nagłówki,
- cytowania RAG-like,
- sekcje bibliografii,
- konstrukcje formalne.

TextMetricsCalculator.cs

- pomocnicze obliczenia metryk tekstowych.

MainForm.TextAudit.cs

- przygotowanie i rekonstrukcja tekstu wklejonego,
- przekazanie treści do TextDetector.

MainForm nie powinien produkować sztucznego końcowego wyniku AI.

53. CZEGO NIE ROBIĆ PRZY KALIBRACJI

Znany tekst testowy może być użyty jako ground truth.

Nie należy jednak ustawiać wyniku na sztywno tylko dlatego, że wiadomo, iż dany tekst pochodzi z ChatGPT, Gemini lub innego modelu.

Prawidłowa procedura:

1. zmierzyć cechy,
2. sprawdzić brakujące sygnały,
3. poprawić metrykę,
4. ponownie przeanalizować,
5. porównać z korpusem tekstów ludzkich i AI.

Celem nie jest dopasowanie jednego przypadku, lecz poprawa generalizacji detektora.

54. PROVIDER ATTRIBUTION DLA TEKSTU

Sam styl tekstu nie pozwala wiarygodnie potwierdzić:

„ten tekst napisał dokładnie ChatGPT”

albo

„ten tekst napisał dokładnie Gemini”.

W przyszłości można rozwijać:

Provider style similarity.

Taki wynik powinien być oznaczony jako podobieństwo stylu, a nie techniczne potwierdzenie providera.

Do kalibracji potrzebny jest oznaczony korpus:

- ChatGPT,
- Gemini,

- Claude,
- teksty ludzkie,
- teksty mieszane.

55. KOLEJNY KIERUNEK – ANALIZA SEGMENTOWA

Jednym z kolejnych kierunków jest analiza tekstu mieszanego człowiek + AI.

Możliwy schemat:

- podział na segmenty około 150-250 słów,
- osobny score dla każdego segmentu,
- porównanie zmian stylu,
- ocena niespójności pomiędzy częściami dokumentu.

Pozwoli to lepiej badać:

- tekst AI po ręcznej redakcji,
- dokument ludzki z fragmentami AI,
- treści składane z kilku źródeł.

56. NAJWAŻNIEJSZE RÓŻNICE 2.0 -> 3.0

W porównaniu z wersją 2.0 dodano lub poprawiono:

TEXT:

- poprawne oddzielenie akapitów od list,
- ostrzejsze rozpoznawanie nagłówków,
- bezpieczniejsze filtrowanie ASCII,
- szablony ofert technicznych,
- równoległość gramatyczną list,
- Syntactic Flatness z analizą sekcyjną i globalnym fallbackiem,
- Transition Frame Signal,
- intro -> lista -> domknięcie,
- cytowania RAG-like,
- wykrywanie bibliografii,
- symetrię nagłówków -> opis,
- echo leksykalne,
- korelację wielu rodzin cech.

CONFIDENCE:

- usunięto praktycznie stałe wartości 42 / 46 / 64,
- Confidence zależy od rzeczywistego score i marginesu od progu,
- uwzględnia ilość materiału,
- uwzględnia liczbę sygnałów,
- wpisy pomocnicze nie używają już prostego „45 + punkty”.

C2PA:

- doprecyzowano oczekiwaną lokalizację c2patool.exe,
- potwierdzono wywołanie C2PA przez ImageDetector przy analizie URL,
- opisano ograniczenie lokalnego Analyze File dla pojedynczych obrazów.

57. PODSUMOWANIE WERSJI 3.0

DetectAI 3.0 rozwija projekt w kierunku wielowarstwowej korelacji Evidence.

Najważniejsze zasady pozostają bez zmian:

- pojedyncza heurystyka tekstowa nie jest technicznym dowodem,
- formalny styl nie jest sygnaturą AI,
- cytowania RAG-like nie potwierdzają same w sobie użycia RAG,
- brak manifestu C2PA nie oznacza braku AI,
- SynthID bez verifera pozostaje Inconclusive,
- NoEvidence nie oznacza „na pewno człowiek”.

DetectAI ma odpowiadać:

- jakie ślady znaleziono,
- skąd pochodzą,
- jak są silne,
- czy potwierdzają się nawzajem,
- jak stabilna jest klasyfikacja,

a nie jedynie zwracać jeden procent „AI”.