PRZEJRZYSTY PROCES

Od zewnętrznych sygnałów do zweryfikowanych decyzji

Widoczny sygnał może uzasadniać sprawdzenie bez przesądzania o podatności, szkodzie ani awarii zabezpieczeń. Utrzymuj ścisłe powiązanie między dowodem, rozumowaniem i decyzją.

> **Obserwacja / Strukturyzacja / Korelacja / Ewaluacja / Weryfikacja / Raport**

01 Obserwacja

Określ usługę, pytanie oceniające i dopuszczalny zakres. Zapisz, co jest widoczne, gdzie zostało znalezione oraz kiedy. Zachowaj datowane odniesienie lub odpowiedni wyciąg dowodowy.

02 Strukturyzacja

Spójnie opisz zasób, źródło, znacznik czasu, istotny fragment oraz ograniczenia. Oddziel bezpośrednie obserwacje od interpretacji oraz brakujących informacji.

03 Korelacja

Porównaj dostępne niezależne źródła. Sprawdź własność, daty i kontekst. Kilka kopii tego samego źródła nie stanowi niezależnego potwierdzenia.

04 Ewaluacja

Wyjaśnij, co potwierdzają dowody, jakie następstwo jest wiarygodne i jakie warunki muszą zajść, aby do niego doszło. Rozważ alternatywne wyjaśnienia.

05 Weryfikacja

Człowiek sprawdza spójność rozumowania, jakość dowodów oraz pozostałą niepewność. Zapisz zdanie odrębne i informacje mogące zmienić ocenę.

06 Raport

Przedstaw ustalenie, znaczenie praktyczne, kolejny krok, osobę odpowiedzialną oraz datę przeglądu. Zachowaj pochodzenie danych wystarczające do odtworzenia wniosków.

> Zaobserwowane: bezpośrednio potwierdzone zapisanym źródłem. Wywnioskowane: interpretacja z jawnym uzasadnieniem. Nieznane: informacja dotąd nieustalona.
>
> Automatyczna ocena punktowa jest tylko sygnałem wejściowym do przeglądu. Zawsze łącz ją z dowodami i ograniczeniami.

LISTA KONTROLNA 01

Ekspozycja zewnętrzna

Dla każdej pozycji zapisz Tak / Nie / Nieznane / Nie dotyczy, źródło potwierdzające oraz właściciela. Traktuj listę jako przewodnik przeglądu, a nie automatyczny wynik zaliczenia.

- ☐ **01 Inwentarz usług**
  Czy wiadomo, które domeny, aplikacje i publiczne interfejsy należą do organizacji?
- ☐ **02 Własność zasobów**
  Czy można odróżnić własne usługi od usług dostawców, porzuconych zasobów i niepowiązanych podobnych domen?
- ☐ **03 Informacje publiczne**
  Czy opublikowane strony i dokumenty ujawniają wyłącznie dane przeznaczone do wiadomości publicznej?
- ☐ **04 Ścieżki konta**
  Czy ścieżki rejestracji, logowania, odzyskiwania dostępu i wsparcia są jasne oraz spójne?
- ☐ **05 Podmioty trzecie**
  Czy zidentyfikowano zewnętrzne usługi wspierające publiczną ścieżkę użytkownika?
- ☐ **06 Informacje o prywatności**
  Czy odwiedzający łatwo znajduje aktualne informacje o przetwarzaniu danych i ścieżkach kontaktu?
- ☐ **07 Zgłaszanie nadużyć**
  Czy działa jasna ścieżka zgłaszania podszywania się, podejrzanych wiadomości i incydentów bezpieczeństwa?
- ☐ **08 Sygnały utrzymania**
  Czy publiczne komunikaty bezpieczeństwa, dokumentacja i informacje o usługach są aktualne?
- ☐ **09 Jakość dowodów**
  Czy podejrzane obserwacje skonfrontowano z prawdopodobnymi łagodnymi wyjaśnieniami?
- ☐ **10 Dalsze kroki**
  Czy każdy istotny punkt ma osobę odpowiedzialną, kolejne działanie i wyznaczoną datę przeglądu?

> **Priorytetyzacja z uwzględnieniem kontekstu**
> Co jest widoczne? Kogo sytuacja może dotyczyć? Jakie dodatkowe warunki są konieczne? Jak mocne są dowody? Jakie działanie ograniczy niepewność lub skutek?

Sama informacja o wersji, dostępna strona lub podejrzana nazwa domeny nie dowodzi wykorzystania podatności ani złośliwego właściciela.

LISTA KONTROLNA 02

AI w edukacji i usługach dla dzieci i młodzieży

Praktyczna rozmowa między właścicielami usług, nauczycielami oraz właściwymi specjalistami. Dla każdego punktu zapisz odpowiedź, dowód potwierdzający oraz nierozstrzygnięte pytania.

01 Cel. Jakie zadanie ma wykonywać AI, a jakie decyzje bezwzględnie pozostają po stronie ludzi?

02 Użytkownicy. Kto ma korzystać z usługi i jak jej projekt odzwierciedla wiek oraz potrzeby odbiorców?

03 Dane. Jakie dane trafiają do systemu? Czy zadanie można sprawdzić na mniej wrażliwych lub syntetycznych przykładach?

04 Cykl życia danych. Czy zasady przechowywania, retencji, usuwania oraz dalszego udostępniania są jasno opisane?

05 Granice działania. Jakich tematów, czynności lub decyzji system powinien odmówić albo przekazać je człowiekowi?

06 Niezawodność. Jak wykrywa się odpowiedzi bez oparcia w źródłach, sprzeczne informacje oraz braki dowodowe?

07 Nadzór człowieka. Kto może przeglądać, korygować, przerwać działanie lub zakwestionować odpowiedź modelu?

08 Komunikacja. Czy użytkownicy wiedzą, że korzystają z AI, i czy rozumieją jej istotne ograniczenia?

09 Dostępność. Czy docelowi użytkownicy rozumieją instrukcje i potrafią skorzystać ze ścieżek wsparcia oraz zgłoszeń?

10 Ewaluacja. Czy zestawy testowe odzwierciedlają docelowe zadania, języki oraz przewidywalne awarie?

11 Reagowanie na incydenty. Kto odbiera zgłoszenia i jakie konkretne kroki następują po otrzymaniu sygnału?

12 Kontrola zmian. Kto zatwierdza istotne modyfikacje i kiedy usługa wymaga ponownej oceny?

> **Przed pilotażem**
> Zapisz dopuszczone zadania, dozwolone dane testowe, osobę weryfikującą, dowody wymagane przed startem oraz warunki natychmiastowego wstrzymania pilotażu.

PRZYKŁAD PRAKTYCZNY

Gdy dowody nie są jeszcze wystarczające

Przykład całkowicie fikcyjny. Nie przedstawia żadnej rzeczywistej usługi, instytucji ani zdarzenia.

> PRZYPADEK SYNTETYCZNY / Niewielka placówka edukacyjna rozważa wdrożenie asystenta AI do zapytań administracyjnych.
>
> Pytanie oceniające: Czy istnieją wystarczające dowody, aby dopuścić ograniczony pilotaż z udziałem danych uczniów?

Dowód	Zapisana obserwacja	Ograniczenie
E-01	Dostarczony opis produktu wskazuje, że przesyłane pytania mogą być zapisywane.	Sam opis nie potwierdza rzeczywistej konfiguracji produkcyjnej.
E-02	Przekazane materiały nie określają żadnego okresu retencji ani terminu usunięcia.	Zabezpieczenia mogą istnieć, lecz dotąd ich nie dostarczono.
E-03	Planowany scenariusz pilotażu zawiera pytania z danymi identyfikującymi uczniów.	Pilotaż nie został jeszcze uruchomiony.

Interpretacja

Dane identyfikujące uczniów mogłyby stworzyć ryzyko przetwarzania danych, którego łatwo uniknąć, dopóki retencja i usuwanie nie są wyjaśnione.

Niewiadome

Rzeczywiste ustawienia retencji, mechanizm usuwania danych, dostęp dostawcy oraz możliwość wyłączenia zapisu historii.

Decyzja

WSTRZYMAJ (HOLD) dla danych identyfikujących uczniów: Podczas wstępnej oceny korzystaj wyłącznie z pytań syntetycznych. Pozostaw dane identyfikujące uczniów poza pilotażem, dopóki właściciel usługi nie dostarczy wystarczających dowodów do zweryfikowanej decyzji.

Kolejne dowody i warunek przeglądu

Poproś o aktualny opis kontroli retencji i usuwania danych, właściwą konfigurację oraz udokumentowany wynik weryfikacji. Oceń ponownie po otrzymaniu dowodów albo przy zmianie celu pilotażu lub zakresu danych.

Wniosek dotyczy wyłącznie kompletności dowodów dla planowanego użycia. Nie przesądza o naruszeniu prawa ani incydencie bezpieczeństwa.

Pochodzenie materiału i dalsza lektura: Opracowano na podstawie opublikowanej metodyki SKOPION, zakresu oceny AI dla dzieci i młodzieży oraz opisu stoiska IGF. Są to materiały edukacyjne bez deklaracji benchmarkowych.

skopion.de/pl/metodyka

skopion.de/pl/igf-2026

Oficjalny profil IGF / Stoisko Village #20 SKOPION Intelligence

ARKUSZ DO WYDRUKU

Karta dowodów i decyzji

Używaj jednej karty dla jednego pytania oceniającego. Dołącz dodatkowe dowody lub notatki, jeśli dostępne miejsce nie wystarcza.

ID karty / data oceny / osoba weryfikująca

Pytanie oceniające / usługa lub zasób / zakres i wyłączenia

ID	Źródło / data	Obserwacja	Ograniczenie

Interpretacja / alternatywne wyjaśnienia / niewiadome lub sprzeczne dowody

Wiarygodne następstwo i warunki konieczne

Poziom pewności (Niski / Średni / Wysoki) i uzasadnienie

Komentarze lub zdanie odrębne po przeglądzie człowieka

Decyzja (w tym HOLD) i uzasadnienie

Działanie / osoba odpowiedzialna / termin docelowy

Dowody wymagane do zamknięcia lub ponownej oceny / data i wyzwalacze zmiany

Wynik po weryfikacji człowieka