CYBERPOLAK

RAPORT BEZPIECZEŃSTWA / FIRMA ALFA

AI używane przez pracowników Firmy Alfa

Analiza ryzyka i perspektywa ISO/IEC 27090:2026

CP-ALFA-AI-01 / WERSJA 2.1
19 PAŹDZIERNIKA 2026

DLA ZARZĄDU FIRMY ALFA

OPRACOWANIE ANALIZY

Adam Jakóbaszek

CyberPolak

Ustalenia techniczne • ryzyko biznesowe • plan działań

Spis treści

Sekcja	Strona
Zakres: AI używane przez pracowników	2
Podsumowanie dla zarządu	3
Dane i wyniki raportu źródłowego	4
Rejestr ustaleń: High i Medium	5
Rejestr ustaleń: Medium i Low	6
Informational i działania hardening	7
Cztery przyczyny architektoniczne	8
Powiązanie z ISO/IEC 27090:2026	9
Plan działań dla Firmy Alfa	10
Metodyka, ograniczenia i źródła	11

Dwa poziomy ustaleń

Ustalenia źródłowe: wyniki opisane w dostarczonym badaniu OpenClaw, z zachowaniem GHSA i statusów raportowanych na 15.09.2026.
Ocena dla Firmy Alfa: konsekwencje dla pracowników korzystających z AI, rekomendowane kontrole i dowody potrzebne do ustalenia, czy dane ryzyko występuje w organizacji.

Zakres: AI używane przez pracowników

Cel oceny

Ocena obejmuje sposób, w jaki pracownicy Firmy Alfa korzystają z AI przy pracy z dokumentami, pocztą, komunikatorami, przeglądarką i narzędziami firmowymi. Celem jest ograniczenie ujawnienia danych, nieuprawnionych działań oraz utraty kontroli nad dostępem delegowanym agentom.

Obszar scope	Co należy objąć oceną
AI w przeglądarce i aplikacjach	Narzędzia zatwierdzone oraz shadow AI; konta służbowe i prywatne; kategorie przekazywanych danych.
Agenci i integracje	MCP, connectors, tool calls, automations, dostęp do plików i możliwość wykonania komend.
Tożsamość i zgody	SSO/MFA, requester identity, role, approval binding, nadawanie i odwoływanie dostępu.
Dane i pamięć	Dokumenty, RAG, uploads, conversations, persistent memory, retencja i content provenance.
Dostawcy i operacje	Warunki wykorzystania danych, logi, alerty, limity, incident response i offboarding.

Granice i nieustalone elementy

Nie dostarczono inwentaryzacji AI, konfiguracji, wersji, logów ani wyników testów z Firmy Alfa. Powyższe obszary określają scope oceny; nie potwierdzają używania konkretnego produktu. Nie zakłada się własnego model training ani wdrożenia OpenClaw.

Testy zewnętrznych usług wymagają uzgodnionego zakresu i uprawnień. Dane do prób powinny być syntetyczne; testowanie nie może polegać na ujawnianiu rzeczywistych danych pracowników lub klientów.

Podsumowanie dla zarządu

Najważniejszy wniosek

Wyniki źródłowego assessment pokazują problem istotny dla organizacji: treść przeczytana przez agenta może wpłynąć na wykorzystanie uprawnień, które mu przyznano. Dokument, wiadomość lub wynik narzędzia nie powinny móc rozszerzyć praw pracownika ani ominąć jego zgody.

Co wynika z dostarczonych danych

Raport źródłowy opisuje 212 granular findings. Po agregacji i weryfikacji threat model przedstawia 27 advisory submissions, osobne zgłoszenie architektoniczne oraz 3 hardening submissions. Nie oznacza to 212 aktywnych podatności u użytkowników ani 212 problemów Firmy Alfa.

Wpływ na Firmę Alfa

W zależności od używanego narzędzia i konfiguracji zagrożone mogą być dokumenty, credentials, prywatność rozmów oraz działania wykonywane przez connectors. Szczególnej uwagi wymagają agenci czytający treści zewnętrzne i jednocześnie posiadający write access, dostęp do host files lub command execution.

Decyzje do podjęcia

1. Zinwentaryzować AI, konta, connectors i kategorie danych.
2. Ograniczyć uprawnienia agentów do konkretnych potrzeb pracownika.
3. Sprawdzić, czy approvals i memory respektują requester identity.
4. Potwierdzić wersje, poprawki i zakres występowania opisanych GHSA.
5. Ustalić owner, termin napraw oraz wymagane evidence i retest.

Status oceny

Analiza dokumentacyjna na podstawie dostarczonego raportu. Występowanie ustaleń w Firmie Alfa: niezweryfikowane. Zgodność z pełnym tekstem ISO/IEC 27090:2026: nieoceniona.

Dane i wyniki raportu źródłowego

Parametr	Dane z raportu [1]
Badany system	OpenClaw, v2026.8.1-beta.2
Commit	baf85e6868e67aebd12ef0e7b77774061fc9db2f
Badanie	10-21 sierpnia 2026; dwóch konsultantów; 3 engineer-weeks.
Raport końcowy	15 września 2026.
Metoda deklarowana w źródle	Static i dynamic testing, dostęp do kodu i dokumentacji, AI-driven processes.

Agregacja wyników

Kategoria	Liczba	Znaczenie
Granular findings	212	Ustalenia przed końcową agregacją.
High / Medium / Low	2 / 16 / 6	24 zgłoszenia w tych klasach severity.
Informational	3	Zgłoszenia defense-in-depth.
Architectural submission	1	Osobne Informational obejmujące 189 outside-scope findings.
Hardening PRs	3	Dodatkowe zgłoszenia bez własnego GHSA.

Ważne: 27 advisory submissions = 24 High/Medium/Low + 3 Informational. Osobne architectural submission daje łącznie 28 identyfikatorów GHSA. Liczby 189 i 212 nie są dodatkowymi zgłoszeniami i nie należy ich sumować z GHSA.

Severity i statusy odtwarzają raport z 15.09.2026. Nie sprawdzano aktualnego stanu zgłoszeń ani poprawek. High GHSA-m969-2g5c-fhwp zamknięto jako not applicable, ponieważ poprawka poprzedziła zgłoszenie.

Rejestr ustaleń: High i Medium

Nazwy techniczne zachowano w języku angielskim. Są to ustalenia badania źródłowego [1], a nie potwierdzone znaleziska w Firmie Alfa.

High severity

GHSA	Ustalenie źródłowe
m969-2g5c-fhwp	System Agent config_get Redaction Bypass. Status: closed / not applicable; fixed before reporting.
9m4p-cqp4-jppq	WhatsApp Login Non-Owner Re-Exposure.

Medium severity / część 1

GHSA	Ustalenie źródłowe
5fwv-rrvp-8xvr	Session Filename Generation Can Execute Tools Denied To The Requesting User
62qm-6fjj-6g23	Restricted User Input Can Reach An Unattended Agent With Write And Command Tools
33m5-gr8m-x3vw	Limited Voice Session Credentials Can Launch An Agent With File And Command Tools
rrxp-5mx8-mvhh	Classic Inbound Voice Calls Inherit The Agent’s Full Tool Authority
qgj5-6x35-9g6f	Config Redaction Exposes Offline Verifiers For Weak Gateway Passwords
h2qg-fxpv-82v6	Model Tools Can Use Credentials From A Disabled Feishu Default Account
m848-fr8q-4vp8	Generic Node Invocation Bypasses Per Agent MCP Tool Denies

Każdy identyfikator w tabelach należy czytać z prefiksem GHSA-. Klasa severity pochodzi ze źródła; zastosowanie w Firmie Alfa wymaga sprawdzenia produktu, wersji i konfiguracji.

Rejestr ustaleń: Medium i Low

Medium severity / część 2

GHSA	Ustalenie źródłowe
w5x7-c87m-3jpc	Requester Scoped Read Denies Do Not Block Outbound Local Attachments
fphf-69cp-h5xw	Telegram Sends Unvalidated Host Files Hidden Behind A Valid Attachment Alias
5rx7-34fw-64qg	Unicode workspaceOnly Parent Fallback Escape
h7jf-rjj9-cww2	MCP App Read Authority Revocation Bypass
v7hh-7676-rg67	Slack download-file Scope Fail-Open
3mq7-q27j-mq7q	Exec Approval CWD Drift
7hw8-jxwc-cwg6	Portal Cookie Origin Collision
85g6-c327-vpf2	Beam Isolated-Profile Catalog Bypass
7jfq-rmfm-29wp	File-Transfer Durable Approval Binding Failures

Low severity

GHSA	Ustalenie źródłowe
5qfh-c63r-7v7g	Curated Memory Provenance Auto-Injection Bypass
chr6-w4m5-57fv	Existing Agents Can Keep Reading Memory After It Is Disabled
6xpv-wwr5-265h	Truncated Tar Bz2 Listings Bypass Skill Extraction Limits
hgv5-f2r3-6v9r	Matrix Case Folding Lets Distinct Accounts Share Authorization
wjfv-5qch-m5vj	Active Memory Can Recall Data For Requesters Denied Memory Tools
g24w-m94m-59qc	Webhook TaskFlow Child Session Cancellation

Informational i działania hardening

Trzy zgłoszenia defense-in-depth

GHSA-ff43-php7-h37v
Incognito Session Authorization Bypass Via Indirect Selectors And Bare Aliases
Zgłoszone jako High; zamknięte jako defense-in-depth hardening po wyjaśnieniu threat model. Same-Gateway incognito/session isolation nie jest wspieraną security boundary.

GHSA-wc97-fpp5-2gjv
Incognito Compat HTTP Session Key Bypass
Zgłoszone jako High; zamknięte jako defense-in-depth hardening po wyjaśnieniu threat model. Compatible HTTP surfaces pozostają wewnątrz trusted-operator Gateway boundary.

GHSA-gfxf-5vv7-cjjj
Conversations List Cross-Agent Account Bridge
Zgłoszone jako High; zamknięte jako defense-in-depth hardening po wyjaśnieniu threat model. W trusted-operator Gateway model nie wskazano lower-trust actor przekraczającego security boundary.

Osobne zgłoszenie architektoniczne

GHSA-2f88-5vh5-3792
Informational: Threat Model Gaps Exposed By Outside-Scope Prompt-Injection And Provenance Findings.
Obejmuje 189 findings pozostających poza aktualnym threat model po feedback od maintainera i ponownej ocenie.

Trzy dodatkowe hardening PRs

github.com/openclaw/openclaw/pull/123846

github.com/openclaw/openclaw/pull/123847

github.com/openclaw/openclaw/pull/123848

Źródło nie przypisuje tym PR własnych GHSA. Nie zweryfikowano ich aktualnego statusu. Nazwy produktów w rejestrze służą identyfikacji ustaleń technicznych, nie stanowią brandingu raportu CyberPolak.

Cztery przyczyny architektoniczne

Poniższe root causes źródło wiąże z 189 outside-scope findings. Są istotne dla doboru narzędzi AI w Firmie Alfa, ale nie dowodzą podatności dowolnego produktu.

1. Agent-global authority

Źródło: treść sterująca agentem może korzystać z jego szerokich connector, browser, memory lub host permissions.
Dla Alfy: e-mail lub dokument nie powinien uruchamiać działań z uprawnieniami przekraczającymi rolę pracownika.
Dowód: policy enforcement wiąże tool call z requester identity i zakresem delegacji.

2. Isolation bez end-to-end security contract

Źródło: sessions, incognito i context filters bywają mechanizmem UX/routing, nie granicą bezpieczeństwa.
Dla Alfy: osobne okno rozmowy nie wystarcza jako ochrona danych działu lub klienta.
Dowód: testy dostępu między kontami i sesjami oraz dokumentacja gwarantowanych granic.

3. Provenance poza policy enforcement

Źródło: pochodzenie lower-trust content nie ogranicza konsekwentnie dostępu do sensitive sinks.
Dla Alfy: pamięć agenta nie powinna nadawać treści zewnętrznej statusu zaufanej instrukcji.
Dowód: zachowanie provenance przez retrieval, memory i kolejne tool calls.

4. Delegation i interaction replay

Źródło: approvals i ponowne interakcje mogą odziedziczyć prawa późniejszego wykonawcy.
Dla Alfy: zgoda pracownika musi dotyczyć konkretnej operacji i jej parametrów.
Dowód: approval binding obejmuje tożsamość, akcję, zasób i ważność; replay jest blokowane.

Powiązanie z ISO/IEC 27090:2026

ISO/IEC 27090:2026 stanowi punkt odniesienia dla zagrożeń bezpieczeństwa AI w cyklu życia. Poniższe powiązanie jest tematyczne; bez pełnej treści normy nie nadajemy numerów postanowień ani ocen „zgodne”.

Obszar oceny	Dane źródłowe	Evidence dla Alfy
Modelowanie zagrożeń	GHSA-2f88-5vh5-3792; cztery root causes.	Inwentaryzacja AI, trust boundaries, dozwolone dane i działania.
Uprawnienia i integracje	MCP tool denies, requester-scoped controls, approval binding.	Role, tool permissions, testy authorization i replay.
Ochrona informacji	Local attachments, memory access, config redaction.	Dane kontrolne, egress logs, zasady memory i kontrola secrets.
Eksploatacja i zmiany	Revocation bypass; pamięć dostępna po wyłączeniu.	Test offboarding, revocation, zmiany konfiguracji i retest.
Wykrywanie i reakcja	Ryzyko wykorzystania connector i host authority.	Audit trail, alerty, procedura odłączenia agenta i incident drill.
Dane treningowe / model	Brak bezpośrednich ustaleń w dostarczonym summary dla tego obszaru.	Zakres po stronie dostawcy; nie wyciągać wniosku o bezpieczeństwie treningu.

Wynik na obecnym etapie

Materiał wskazuje obszary, które należy ocenić i udokumentować. Nie stanowi pełnej oceny ISO/IEC 27090:2026 ani certyfikacji. Dla każdego właściwego postanowienia należy później zapisać: zastosowanie, evidence ID, wynik, lukę, działanie i residual risk.

Plan działań dla Firmy Alfa

Kolejność jest rekomendacją CyberPolak wynikającą z analizy źródła. Nie zastępuje oceny severity i likelihood w rzeczywistym środowisku Firmy Alfa.

Kiedy	Działanie	Odbiór / owner
Najpierw	Spis AI, accounts, wersji, connectors, praw i kategorii danych.	Rejestr zatwierdzony przez IT i właścicieli biznesowych.
Przed rozszerzeniem uprawnień	Ograniczenie write access, command execution i dostępu do secrets; sprawdzenie poprawek dla właściwych GHSA.	Potwierdzone konfiguracje i wersje; IT / bezpieczeństwo.
Przed dostępem do danych wrażliwych	Testy authorization, memory isolation, egress i indirect prompt injection.	Udokumentowane wyniki; zespół bezpieczeństwa.
Przed odbiorem	Ćwiczenie revocation, offboarding i awaryjnego odłączenia integracji.	Logi i potwierdzenie cofnięcia dostępu; operacje IT.
Po zmianach	Retest, aktualizacja ryzyka i decyzja dotycząca residual risk.	Właściciel systemu i właściciel ryzyka.

Karta weryfikacji ustalenia

Referencja: GHSA-h7jf-rjj9-cww2 / MCP App Read Authority Revocation Bypass.
Pytanie dla Alfy: czy odwołanie uprawnienia odbiera odczyt również aktywnemu agentowi?
Warunek zastosowania: używanie odpowiadającej integracji i wersji.
Evidence: konfiguracja przed/po revocation, test odczytu, audit log.
Status lokalny: do weryfikacji.
Owner / termin: do wskazania.
Retest / residual risk: do oceny po uzyskaniu wyników.

Metodyka, ograniczenia i źródła

Podstawa opracowania

[1] OpenClaw, Patch the Planet - Security Assessment (Summary Report), 15.09.2026; plik przekazany przez zamawiającego: 2026-09-openai-openclaw-securityreview.pdf. Autor wskazany w dokumencie: Trail of Bits; zleceniodawca wskazany w dokumencie: OpenAI. Dane liczbowe i wyniki pochodzą z tego materiału, a nie z testów wykonanych przez CyberPolak w Firmie Alfa.

Wykorzystane części: Project Targets (s. 3), Executive Summary (s. 4-5), Summary of Advisory Submissions (s. 6-8). Numeracja odpowiada stopkom źródła, nie numerom stron przeglądarki PDF.

[2] ISO/IEC 27090:2026, publiczny opis katalogowy ISO:
https://www.iso.org/standard/27090
Status publikacji sprawdzono 19.10.2026. Pełnego tekstu normy nie analizowano.

Wkład CyberPolak

CyberPolak opracował polską analizę zastosowania ustaleń do scope AI używanego przez pracowników Firmy Alfa, rekomendacje, plan weryfikacji oraz powiązanie tematyczne z ISO. Zachowano angielskie nazwy techniczne i identyfikatory źródłowe. Logotypów źródłowych nie wykorzystano.

Czego nie sprawdzono

Nie potwierdzono używanych w Alfie produktów, wersji, konfiguracji ani występowania opisanych podatności. Nie przeprowadzono testów, przeglądu logów ani retest. Nie weryfikowano aktualnego statusu GHSA i PR. Klasy severity odpowiadają raportowi źródłowemu i nie są automatycznie oceną ryzyka w Alfie.

Nie oceniono zgodności z NIS2, DORA ani AI Act. Stwierdzenie braku zastosowania konkretnego GHSA wymaga uzasadnienia. Brak danych lub brak testu należy zapisać jako „niezweryfikowane”, nie jako „bezpieczne”.