DarhimLabs

Zacznij trial →

SECURITY WERSJA V2.1 · OBOWIĄZUJE OD 01 MAJA 2026

· OSTATNIA AKTUALIZACJA 22 KWIETNIA 2026

Bezpieczeństwo DarhimLabs

Pobierz PDF Zobacz DPA

SPIS TREŚCI

Bezpieczeństwo DarhimLabs jest projektowane jako część produktu, a nie dodatkowa warstwa po wdrożeniu. Platforma obsługuje rozmowy klientów, konfiguracje agentów AI, dokumenty bazy wiedzy, webhooki, klucze API, transkrypcje Voice i dane compliance, dlatego każdy moduł musi mieć kontrolę dostępu, audyt, retencję, izolację tenantów i bezpieczne domyślne ustawienia. Ten dokument opisuje architekturę, proces reagowania na incydenty, program zgłaszania podatności i sposób prowadzenia testów bezpieczeństwa.

1. Model bezpieczeństwa

DarhimLabs stosuje model
aplikacji, bazie danych, sto
operacyjnych. Każda warst
kontrole są powielane. Przy
dane workspace'ów są doc
PostgreSQL. Mutacje są wa
logu.

Zasada minimalnych upraw
zespołu DarhimLabs. Role
działania administracyjne w
produkcyjny po stronie Dar

przeglądany i dostępny wyłącznie osobom, które potrzebują go do obsługi incydentu, utrzymania systemu, migracji lub zgłoszenia wsparcia. Bezpieczeństwo obejmuje również AI. Modele językowe są traktowane jak komponenty wysokiego ryzyka, ponieważ mogą generować treści, wykonywać narzędzia, klasyfikować dane i korzystać z dokumentów RAG. Dlatego DarhimLabs stosuje prompt-injection checks, polityki zatwierdzeń, ograniczenia domen, maskowanie PII, kosztowe limity wykonania i rejestrację tool calls. Wrażliwe akcje, takie jak płatności lub zmiany danych w CRM, mogą wymagać Approval Inbox.

RLS + RBAC

Izolacja workspace'ów i role per moduł.

TLS 1.3 + at-rest

Szyfrowanie transmisji i danych w storage.

Monitoring 24/7

Alerty, SIEM, Sentry i health checks.

Audit log

Rozliczalność mutacji i akcj

2. Architektura bezpieczeń
Architektura DarhimLabs je
przechodzi przez ochronę
aplikacyjna Next.js odpowi
nagłówki bezpieczeństwa i
z RLS, storage, auth i realti
vault, rotację kluczy i ograniczenie dostępu do tokenów integracji. Monitoring

zbiera logi aplikacyjne, błędy, zdarzenia bezpieczeństwa, health checks i alerty compliance.
Każda warstwa ma odrębne mechanizmy kontroli. Edge blokuje ruch automatyczny, reguły WAF i anomalie. Aplikacja waliduje Zod schema, sprawdza workspace context i rate limit. Baza danych egzekwuje izolację tenantów przez polityki RLS, indeksy i foreign keys. Monitoring pozwala wykrywać regresje, błędy CSP, nadużycia API, nietypowe logowania oraz próby omijania limitów.

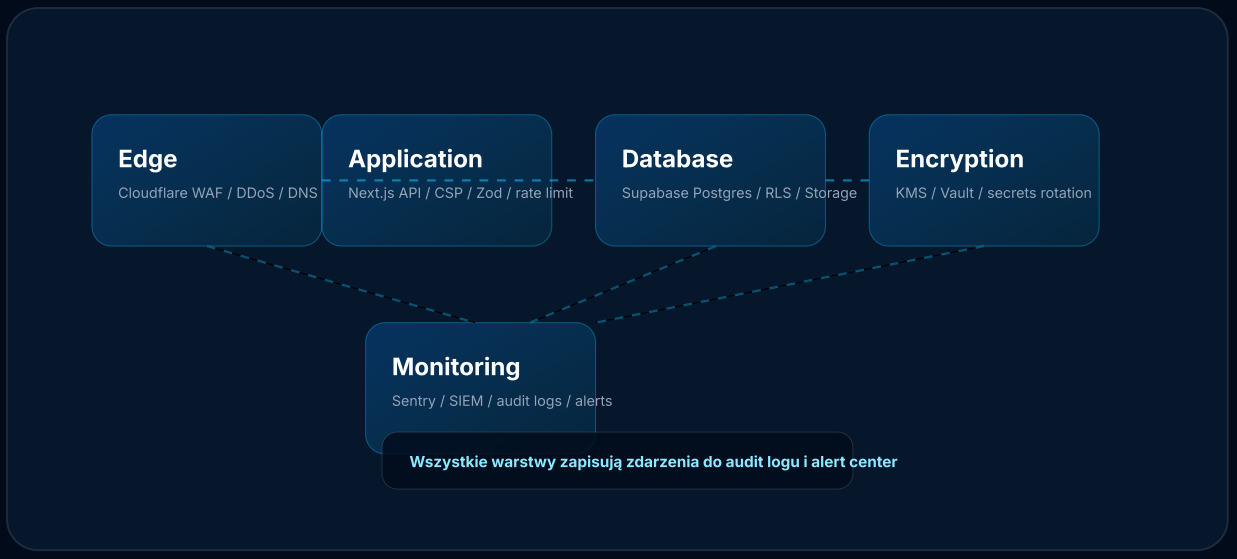

Diagram jest uproszczeniem. Wdrożenie produkcyjne obejmuje też kolejki, webhook retries, storage plików, integracje OAuth, API v2, system powiadomień push i workerów asynchronicznych. Dla każdego komponentu stosujemy osobne sekrety, rate limits, walidację wejścia, logowanie błędów i ograniczony zakres danych w payloadach.
3. Dane, szyfrowanie i kontrola dostępu
Dane w tranzycie są chronione przez TLS 1.3. Dane w spoczynku korzystają z zabezpieczeń dostawców infrastruktury oraz dodatkowych mechanizmów aplikacyjnych dla sekretów i tokenów. Klucze API, tokeny OAuth, sekrety webhooków i credentials in
Dostęp do nich jest ogranic
Kontrola dostępu działa na
workspace, uprawnienia m
danych. Dzięki temu nawet
odczytać danych innego w
dodatkowe mechanizmy: S
eksport audit logów i polity
Backupy są tworzone regu
okresowo. Dane usuwane

zgodnie z harmonogramem retencji. Szczegółowe zasady zwrotu i usuwania danych opisuje DPA, a prawa osób opisuje przewodnik RODO.

Kontrole danych i dostępu

KONTROLA
TLS 1.3

ZAKRES
Ruch przeglądarka, API, webhooki

CEL
Poufność i integralność transmisji

KONTROLA
RLS

ZAKRES
Postgres i workspace_id

CEL
Izolacja tenantów na poziomie danych

KONTROLA
MFA

ZAKRES
Konta zespołu i Enterprise

CEL
Ograniczenie przejęcia k

KONTROLA
Audit log

ZAKRES
Mutacje, role, sekrety, in

CEL

Rozliczalność i forensic readiness

4. AI, compliance i izolacja tenantów

DarhimLabs wykorzystuje modele językowe do generowania odpowiedzi, klasyfikacji intencji, tworzenia draftów, RAG, benchmarków jakości i automatyzacji. Dane klientów nie są używane do trenowania modeli DarhimLabs ani modeli dostawców LLM w domyślnej konfiguracji. Wywołania AI są ograniczane do danych potrzebnych dla konkretnej operacji, a payloady mogą być redagowane przez mechanizmy PII masking.

Izolacja tenantów dotyczy także RAG. Dokumenty, chunki, embeddings i cytowania są przypisane do workspace'u oraz źródła. Zapytania testowe i produkcyjne nie powinny mieszać źródeł między klientami. Dodatkowe kontrole obejmują citation coverage, freshness, quality score, alerty regresji i możliwość wyłączenia demo data poza trybem demo.

Moduły AI Agents i Bot Builder mają polityki zatwierdzeń dla działań wysokiego ryzyka. Bot może sugerować akcję, ale wykonanie może wymagać człowieka, zwłaszcza przy płatnościach, zmianach danych CRM, wysyłce SMS, rezerwacjach, escalations lub działaniach w systemach zewnętrznych. Każde wywołanie narzędzia zapisuje parametry, wynik, czas i koszt tam, gdzie jest to potrzebne do audytu.

Prompt injection checks ograniczają wykonywanie instrukcji pochodzących z niezaufanych źródeł.

Approval Inbox pozwala zatrzymać akcję AI przed wykonaniem w systemie zewnętrznym.

Quality Lab wykrywa regresje jakości, halucynacje i spadek citation coverage.

Model routing może rozdzielać zadania według jakości, kosztu, latency i wymogów compliance.

5. Cykl życia incydentu

Incident response w DarhimLabs ma pięć etapów: detection, triage, containment, eradication, recovery i post-mortem. Detection oznacza wykrycie sygnału przez monitoring, a
badacza lub logi dostawcy.
dotknięte moduły i właścici
przykład przez wyłączenie
Eradication usuwa przyczy
RLS, zależność lub proces.
incydent nie wraca. Post-m
wpływ na dane, komunikac
poważnych zdarzeń przyg
jeżeli incydent dotyczył ich

Jeśli incydent stanowi naruszenie ochrony danych osobowych, uruchamiamy procedurę RODO: IODO ocenia ryzyko, a DarhimLabs informuje Administratora bez zbędnej zwłoki. Jeżeli DarhimLabs jest Administratorem i naruszenie wymaga zgłoszenia do PUODO, stosujemy termin 72 godzin od stwierdzenia naruszenia zgodnie z art. 33 RODO. Szczegóły są spójne z DPA i RODO.

Severity i czasy reakcji bezpieczeństwa

SEVERITY
P1 krytyczny

PRZYKŁAD
Potwierdzony dostęp do danych wielu workspace'ów lub sekretów produkcyjnych

LEAD TIME
do 15 minut

REAKCJA
SRE on-call, Security Lead, IODO, status page i dedykowany kanał klienta

SEVERITY
P2 wysoki

PRZYKŁAD
Błąd izolacji tenantów ograniczony do jednego workspace'u albo ryzyko ujawnienia PII

LEAD TIME
do 1 godziny

REAKCJA
Security triage, owner w

SEVERITY
P3 średni

PRZYKŁAD
Podatność o ograniczon
danych

LEAD TIME

do 4 godzin

REAKCJA

Ticket bezpieczeństwa, patch w normalnym cyklu hotfix

SEVERITY

P4 niski

PRZYKŁAD

Twarde nagłówki, kosmetyka CSP, niski wpływ na dane

LEAD TIME

do 1 dnia roboczego

REAKCJA

Backlog security hardening i kwartalny przegląd

6. Bug bounty program

DarhimLabs przyjmuje zgłoszenia podatności od badaczy bezpieczeństwa i klientów. Program bug bounty obejmuje publiczne powierzchnie produkcyjne, w szczególności app.darhimlabs.pl, api.darhimlabs.pl, widget klienta, endpointy API v1/v2, mechanizmy auth, izolację tenantów, webhooki, widget embed i panel dashboard. Najwyższy priorytet mają podatności pozwalające na nieautoryzowany dostęp do danych, obejście RLS, eskalację uprawnień, przejęcie konta, wyciek sekretów lub wykonanie akcji w cudzym workspace. Poza zakresem są środowiska prywatne dev/staging, systemy osób trzecich niezależne od DarhimLabs, ataki DDoS, social engineering, spam, fizyczne ataki na pracowników, masowe skanowanie bez ograniczeń, raporty bez wpływu bezpieczeństwa i znane problemy w bibliotekach, jeżeli nie da się ich wykorzystać w DarhimLab
masowo danych, nie mody
usługi i nie ujawniać podat
Nagrody są oceniane wedł
zakresu. Orientacyjny zakr
krytyczne podatności umo
wymaga raportu zawierają
payload i rekomendację na
prowadzonych zgodnie z z

In scope

App, API, widget, auth, RLS, tenant isolation.

Out of scope

DDoS, social engineering, spam, third-party.

Safe harbor

Odpowiedzialne testy bez naruszania danych.

7. Testy penetracyjne

DarhimLabs prowadzi zewnętrzne testy penetracyjne co najmniej raz w roku oraz po istotnych zmianach architektury, takich jak nowy model auth, nowe API, publiczny widget, zmiana izolacji tenantów albo wdrożenie funkcji wysokiego ryzyka. Zakres obejmuje OWASP Top 10, OWASP API Security Top 10, kontrolę tenant isolation, SSRF, XSS, CSRF, IDOR, błędy autoryzacji, bezpieczeństwo webhooków i konfigurację nagłówków.

Preferujemy niezależnych dostawców z doświadczeniem w aplikacjach SaaS, API i cloud security, na przykład Securitum lub równoważnych vendorów. Testy obejmują środowisko kontrolowane, przygotowany zakres, konto testowe, ograniczenia czasowe i kanał awaryjny. Wyniki trafiają do rejestru ryzyk, a naprawy są priorytetyzowane według CVSS, wpływu na klientów i możliwości nadużycia.

Executive summary testów
Enterprise po NDA. Pełny r
payloady i konfigurację sys
zakresie, jeżeli jest to uzas
innych klientów. Status dzi
zarządzania ryzykiem.

Testy roczne obejmują apli
Testy po dużych zmianach
architektury.

Krytyczne i wysokie podatności mają priorytet hotfix lub release blokujący.

Podsumowania audytów są dostępne dla Enterprise po NDA.

8. Vulnerability disclosure

Podatności bezpieczeństwa należy zgłaszać na adres security@darhimlabs.pl. W temacie wpisz „Security disclosure" oraz krótki opis wpływu. W raporcie umieść kroki reprodukcji, adres URL, payload, oczekiwany i rzeczywisty wynik, zrzuty ekranu lub logi, ocenę wpływu oraz informację, czy podatność mogła dotknąć danych osobowych. Nie wysyłaj haseł, pełnych tokenów, danych kart płatniczych ani danych osób trzecich.

Potwierdzamy otrzymanie raportu w ciągu 24 godzin roboczych dla zgłoszeń o potencjalnie wysokim wpływie. Następnie wykonujemy triage, klasyfikację severity i komunikujemy plan naprawy. Oczekujemy odpowiedzialnego ujawnienia przez 90 dni lub do czasu uzgodnionego przez strony. Jeżeli podatność jest krytyczna i aktywnie wykorzystywana, możemy poprosić o dłuższe embargo do czasu ochrony klientów.

DarhimLabs może wnioskować o CVE dla kluczowych podatności w komponentach publicznych lub bibliotekach, jeżeli spełniają kryteria ekosystemu. Dla podatności w samej aplikacji SaaS priorytetem jest szybka naprawa, komunikacja z klientami i ograniczenie ryzyka, a nie publiczny rozgłos. Badacze działający w dobrej wierze są objęci safe harbor i nie będą ścigani za testy zgodne z zasadami programu.

PGP public key

Jeśli raport zawiera wrażliwe szczegóły, zaszyfruj wiadomość kluczem PGP. Pełny klucz publikujemy w security.txt i na żądanie przez e-mail.

Zgłoś podatność →

```
-----BEGIN PGP PUBLIC KEY
Version: DarhimLabs Secu

mDMEZ0DSECURITYDARHIMLABS
Fingerprint: 7F4A 2026 1
Contact: security@darhim
-----END PGP PUBLIC KEY
```

Pliki cookies

KONTROLA PRYWATNOSCI

Uzywamy cookies niezbednych do dzialania strony. Analytics i marketing wlaczymy dopiero po zgodzie. Szczegoly znajdziesz w **polityce cookies**.

Dodatkowe informacje o zgodności, transferach i przetwarzaniu danych znajdują się w DPA, Polityce prywatnościoraz SLA. Status działania platformy publikujemy na stronie /status.

DarhimLabs

AI Business OS dla rozmow, leadow, agentow
AI, voice, quality lab i compliance.

DarhimLabs Sp. z o.o.
ul. Pilsudskiego 1 m. 43
09-200 Sierpc
kontakt@darhimlabs.pl

Produkt

Platforma
AI Agents
Voice
Compliance
Integracje
API v2
Cennik

Zasoby

Blog
Case studies
Roadmap
Changelog
Docs
Status

Pliki cookies
KONTROLA PRYWATNOSCI

Uzywamy cookies niezbednych do dzialania strony. Analytics i marketing wlaczymy dopiero po zgodzie. Szczegoly znajdziesz w **polityce cookies**.

Firma

O nas

Kariera

Partnerzy

Kontakt

Press kit

Zgodność

Polityka prywatnosci

Regulamin

RODO

DPA

Bezpieczeństwo

Cookies

NIP: do uzupelnienia REGON: do uzupelnienia

SOC2 ISO27001 GDPR

PL Status: degraded

Pliki cookies

KONTROLA PRYWATNOSCI

Uzywamy cookies niezbednych do dzialania strony. Analytics i marketing wlaczymy dopiero po zgodzie. Szczegoly znajdziesz w **polityce cookies**.