Inteca | IT's about business

Security Overview

PractIQ - AI Delivery Operating System

Inteca | Red Hat

Wprowadzenie

Wdrożenie AI w środowiskach enterprise stawia nowe wymagania w zakresie bezpieczeństwa. Agenci AI operują na danych organizacyjnych, wywołują zewnętrzne modele językowe i podejmują autonomiczne działania w imieniu użytkowników — wprowadzając nowe klasy ryzyka, dla których tradycyjne mechanizmy kontroli dostępu nie zostały zaprojektowane.

PractIQ to rozwiązanie, które zostało zaprojektowane z myślą o środowiskach enterprise o wysokich wymaganiach bezpieczeństwa — w szczególności sektora finansowego, bankowego i ubezpieczeniowego. Bezpieczeństwo nie jest w PractIQ warstwą dodaną na końcu — jest wbudowane w architekturę platformy na każdym poziomie: od izolacji środowisk wykonawczych, przez kontrolę dostępu do danych, po ochronę przed niekontrolowanym przepływem informacji do modeli językowych.

Niniejszy dokument opisuje kluczowe zasady i mechanizmy bezpieczeństwa, na których opiera się platforma PractIQ.

Kluczowe zasady bezpieczeństwa

1. Izolacja środowisk wykonawczych

Każda praktyka uruchamiana w PractIQ działa w oddzielnym, ściśle kontrolowanym środowisku kontenerowym na platformie Kubernetes — w rekomendowanej konfiguracji produkcyjnej na Red Hat OpenShift. Środowisko każdej praktyki jest predefiniowane i niemodyfikowalne w trakcie działania: agent ma dostęp wyłącznie do narzędzi, danych i usług niezbędnych do realizacji konkretnej praktyki. Żadne zasoby spoza zdefiniowanego zakresu nie są dostępne w środowisku pracy.

Izolacja środowisk eliminuje ryzyko nieautoryzowanego dostępu między praktykami oraz zapewnia pełną przewidywalność działania agentów — niezależnie od tego, ilu użytkowników i agentów działa równocześnie na platformie.

Inteca Sp. z o.o.
ul. Powstańców Śląskich 9
53-332 Wrocław
+48 71 715 60 91
biuro@inteca.pl
inteca.com
practiq.tech

2. Zero Trust i kontrola dostępu oparta na atrybutach

Dostęp do danych — zarówno dla użytkowników, jak i agentów AI — oparty jest na modelu Zero Trust z kontrolą dostępu opartą na atrybutach (ABAC — Attribute-Based Access Control). Oznacza to, że uprawnienia nie są przyznawane na podstawie samej roli użytkownika, lecz na podstawie kombinacji atrybutów: tożsamości, kontekstu wykonywanej praktyki, klasyfikacji danych i bieżącego stanu sesji.

W praktyce: agent AI wykonujący praktykę analizy wymagań ma dostęp wyłącznie do danych i narzędzi zdefiniowanych dla tej praktyki — nie może uzyskać dostępu do danych powiązanych z inną praktyką ani wywołać usług spoza swojego zakresu uprawnień. Podejście to jest niezależne od metody pobierania danych — obowiązuje zarówno dla wywołań przez protokół MCP, jak i bezpośrednich wywołań usług.

3. Brama LLM i guardrails

Wszystkie wywołania modeli językowych przechodzą przez centralną bramę LLM opartą na LiteLLM, wyposażoną w zestaw mechanizmów kontrolnych (pre-call hooks) uruchamianych przed każdym wysłaniem danych do modelu. Mechanizmy te skanują i walidują dane wejściowe, zapewniając że żadne informacje niespełniające zdefiniowanych kryteriów bezpieczeństwa nie trafią do zewnętrznego lub lokalnego modelu językowego.

Brama LLM stanowi centralny punkt kontroli i audytu wszystkich interakcji agentów z modelami językowymi — umożliwiając pełne logowanie, monitoring i wykrywanie anomalii.

4. Lokalna klasyfikacja danych i ochrona przed wyciekiem (DLP)

PractIQ implementuje lokalny mechanizm klasyfikacji i ochrony danych (Data Loss Prevention) oparty na ontologii dziedzinowej — nie na wzorcach tekstowych. Każda praktyka operuje wyłącznie na danych zdefiniowanych w ontologii organizacyjnej. Dane w grafie wiedzy są klasyfikowane według poziomu wrażliwości, a lokalny gateway rozumie tę klasyfikację i automatycznie maskuje wrażliwe dane przed ich wysłaniem do modelu językowego.

Podejście oparte na ontologii — w odróżnieniu od klasycznych systemów DLP opartych na wyrażeniach regularnych — zapewnia spójne i kontekstowe maskowanie danych między wywołaniami, eliminuje fałszywe alarmy wynikające z dopasowań wzorców oraz działa lokalnie, bez narzutu na bramę LLM i bez konieczności wysyłania danych do zewnętrznych systemów klasyfikacji.

Inteca Sp. z o.o.
ul. Powstańców Śląskich 9
53-332 Wrocław
+48 71 715 60 91
biuro@inteca.pl
inteca.com
practiq.tech

Bezpieczeństwo w kontekście Red Hat OpenShift

Rekomendowanym środowiskiem produkcyjnym dla PractIQ jest Red Hat OpenShift, który dostarcza dodatkowe warstwy bezpieczeństwa na poziomie infrastruktury. Wbudowane mechanizmy RBAC (Role-Based Access Control) i network policies zapewniają izolację sieciową między komponentami platformy. Automatyczne skanowanie obrazów kontenerowych wykrywa znane podatności przed wdrożeniem. Audit logging na poziomie klastra zapewnia pełną ścieżkę audytu dla wszystkich operacji. Integracja z Red Hat Build of Keycloak umożliwia zarządzanie tożsamością i dostępem zgodne ze standardami enterprise (SSO, MFA, federacja tożsamości).

Wdrożenie PractIQ on-premise w klastrze OpenShift — rekomendowane dla organizacji z najwyższymi wymaganiami bezpieczeństwa, takich jak banki i instytucje finansowe — zapewnia pełną suwerenność danych: żadne dane organizacyjne nie opuszczają infrastruktury klienta.

Podsumowanie

Podejście PractIQ do bezpieczeństwa opiera się na czterech wzajemnie wzmacniających się zasadach: izolacji środowisk wykonawczych, kontroli dostępu Zero Trust opartej na atrybutach, centralnej bramie LLM z mechanizmami guardrails oraz lokalnej klasyfikacji i ochronie danych opartej na ontologii. W połączeniu z mechanizmami bezpieczeństwa Red Hat OpenShift tworzą one wielowarstwową architekturę bezpieczeństwa, która zapewnia że agenci AI działają wyłącznie w zdefiniowanych granicach — bez możliwości nieautoryzowanego dostępu do danych, niekontrolowanego wywołania modeli językowych czy wycieku wrażliwych informacji.

Więcej informacji: inteca.com | practiq.tech

Inteca Sp. z o.o.
ul. Powstańców Śląskich 9
53-332 Wrocław
+48 71 715 60 91
biuro@inteca.pl
inteca.com
practiq.tech