Projektni zadatak – Secure Chatbot-as-a-Service

1. Opis arhitekture

Arhitektura sustava "Secure Chatbot-as-a-Service" temelji se na višeslojnom web rješenju:
- Frontend aplikacija u Next.js (TypeScript) i služi kao SPA korisničko sučelje.
- API Gateway (FastAPI) prosljeđuje sve zahtjeve prema autentikacijskom servisu, servisima za chatbotove i zapisivanju logova.
- Autentifikacija i sesije obrađuju se putem Auth servisa, koristeći JWT i Redis za pohranu sesija.
- Chatbot servis komunicira s OpenAI API-jem ili lokalnim Llama.cpp back-endom.
- Audit Log servis bilježi sve zahtjeve korisnika.
- Background worker (Celery + Redis) obrađuje asinkrone zadatke poput nadzora korištenja resursa i filtriranja promptova.
- Baza podataka je PostgreSQL s pgvector ekstenzijom, uz tenant segregaciju podataka po korisniku.
- Nginx reverse proxy s AI-WAF slojem filtrira ulazni promet.
- CI/CD pipeline koristi GitHub Actions za automatizirani deployment i sigurnosne provjere (Dependabot, Trivy).

2. Funkcionalni zahtjevi

ID	Opis	Status
FR-1	Registracija / prijava (korisničko ime + lozinka, Argon2 hash)	Obavezno
FR-2	Stvaranje, uređivanje i brisanje chatbot instanci	Obavezno
FR-3	Čuvanje i prikaz povijesti razgovora	Obavezno
FR-4	Konfigurabilni prompt filter (Regex + LLM self-check)	Pojačani
FR-5	Rate-limiting i statistika poziva prema OpenAI API-ju	Pojačani
FR-6	Audit log svih zahtjeva/odgovora za forenziku	Pojačani
FR-7	Izvoz korisničkih podataka (GDPR "Right-to-access")	Dodatno

3. Model prijetnji (STRIDE)

Kategorija	Primjer prijetnje	Mitigacija
S – Spoofing	Probijanje lozinke i krađa tokena	MFA, Argon2, HTTP-Only cookies
T – Tampering	Manipulacija prompta između UI-ja i back-enda	HTTPS, HMAC potpisi poruka
R – Repudiation	Poricanje zlonamjernih zahtjeva	Neizmjenjivi audit log (WORM)
I – Information Disclosure	Chat history bleed između tenant-ova	Tenant ID izolacija, row-level security
D – Denial of Service	Prekomjerni API pozivi	Global + per-user rate-limit, circuit breaker
E – Elevation of Privilege	Prompt injection za dobivanje admin pristupa	LLM output sanitization + RBAC provjere

4. Plan penetracijskog testiranja

Korak	Alat/Metoda	Cilj
1	OWASP ZAP Baseline Scan	Brzi pregled (XSS, CSRF, Headers)
2	Burp Suite Intruder	Fuzzing endpointa za prompt injection
3	Custom LLM Payload Generator	Test bypassa prompt filtra
4	sqlmap	Verifikacija RLS postavki
5	Locust	DoS scenariji + stres OpenAI kvota
6	Ručna provjera RBAC	Pokušaj eskalacije tenant granica