LimaChat v3.1

AI Agent za prodaju i tehničku podršku

Kompletna specifikacija funkcionalnosti

Ovaj dokument sadrži kompletan popis funkcionalnosti LimaChat sistema. Bez marketinga, bez ulepšavanja — samo činjenice o tome šta sistem radi, kako radi i šta možete da očekujete.

1. AI prodajni agent

LimaChat preuzima ulogu prvog kontakta za sve prodajne upite. Razume šta korisnik traži, pronalazi tačne informacije iz baze znanja i vodi razgovor do narudžbine.

Funkcionalnost	Detalji
Prepoznavanje namere	Automatski detektuje da li korisnik pita o proizvodima, cenama, paketima ili želi da naruči. Radi u 4 sloja: sticky session → regex brzi put → keyword matching → LLM klasifikacija za nejasne slučajeve.
Baza znanja	Crawluje vaš sajt, parsira sadržaj, klasifikuje stranice (pricing, FAQ, blog, procedura) i gradi strukturovane Q&A atome. Kad korisnik pita — traži odgovor u 5 nivoa: Tier-0 (obavezna pravila) → Tier-0.5 (katalog) → Tier-0.8 (Q&A parovi) → Tier-1 (crawlovane stranice) → Tier-2 (Brave web search).
Zaključavanje paketa	Kad korisnik kaže koji paket želi (npr. "hoću Premium"), sistem to zaključava — AI ne sme da nudi skuplji paket, ne sme da menja izbor, ne sme da upsell-uje.
Prikupljanje podataka	Kada korisnik odluči da naruči, AI traži ime, email i telefon — jedno po jedno. Podaci se izvlače regex-om direktno iz poruka korisnika (ne iz AI odgovora) — nema halucinacije.
Potvrda narudžbine	AI prikazuje prikupljene podatke i traži potvrdu. Kad korisnik kaže "da" — sistem to detektuje na dva nivoa: semantika (AI prepoznao) + code-level interceptor (regex fallback). Garantovano 100% fire rate.
Email notifikacija	Potvrđena narudžbina se šalje na email prodajnog tima sa svim podacima korisnika + kompletnom transkriptom razgovora + geo lokacijom korisnika.
Praćenje cenovnika	Ako se konfiguriše URL cenovnika, sistem proverava svaka 2 minuta da li su se cene promenile (SHA-256 hash). Ako jesu — AI koristi live cene, a KB se automatski rebuild-uje u pozadini.
CTA (poziv na akciju)	Na kraju svakog prodajnog odgovora može se dodati konfigurabilan poziv na akciju (npr. link ka naručivanju). Uključuje/isključuje se iz admin panela.

2. AI agent za tehničku podršku

Odgovara na tehnička pitanja korisnika koristeći bazu znanja napravljenu od sadržaja sa vašeg sajta, help centra i dokumentacije.

Funkcionalnost	Detalji
Tehnički odgovori	Pretražuje bazu znanja za relevantne odgovore. Ako informacija ne postoji u bazi — kaže korisniku da proveri sajt ili kontaktira podršku. Nikada ne izmišlja odgovore.
Status page integracija	Automatski proverava status page (Atlassian Statuspage, Cachet, RSS, HTML) pri svakom support upitu. Ako postoji aktivan incident — obaveštava korisnika pre nego što počne da traži u bazi.
Eskalacija ka živom agentu	Korisnik može da zatraži kontakt sa živom podrškom. AI prikuplja ime, email i opis problema, pa prosledjuje support timu na email sa kompletnim transkriptom.
Symptom pattern matching	Konfigurabilan skup obrazaca simptoma (sajt ne radi, email problem, DNS, SSL, backup, WordPress, IP blokada, transfer domena). Kada korisnik opiše simptom, sistem odmah zna u kojoj oblasti da traži.
Multi-intent	Ako korisnik u jednoj poruci pita i za cenu i za tehničko pitanje ("Koliko košta hosting i kako da podesim DNS?"), sistem detektuje oba intenta i daje odgovor koji pokriva oba dela.
Web search fallback	Kad lokalna baza nema odgovor, sistem koristi Brave Search API da pronađe relevantan sadržaj sa interneta. Konfigurisano po klijentu.

3. Jezici i lokalizacija

Funkcionalnost	Detalji
Jezici	Srpski i engleski. Automatska detekcija u 9 slojeva: dijaktitika → ćirilica → latinične reči → gramatički obrasci → kratke poruke. Nikad ne pogrešno detektuje tehničke termine (DNS, VPS) kao engleski.
Session stickiness	Jednom kad se ustanovi jezik razgovora, ne menja se osim ako korisnik eksplicitno predje na drugi jezik (minimum 2 jake reči na novom jeziku).
Srpski dijalekt	Ekavica + latinica — uvek. Prompt instrukcija + post-processing enforcement. Ako AI napiše "gdje" ili ćirilicu, sistem automatski ispravlja u "gde" / latinicu.
Persona po jeziku	Odvojene persone za srpski i engleski. Svaka se može konfigurisati nezavisno iz admin panela.

4. Admin panel

Kompletna kontrola nad AI agentom bez ikakvog programiranja.

Funkcionalnost	Detalji
Dashboard	Pregled aktivnosti: broj razgovora, intenti, jezici, prosek ocena, geolocija korisnika, grafikon aktivnosti.
Konfiguracija AI-a	Persona (SR/EN), pozdravna poruka, oproštajna poruka, CTA tekstovi, max dužina odgovora (sales/support/apsolutni), prikupljanje podataka — sve se menja u admin panelu.
Widget dizajner	Boje, pozicija, avatar, ime agenta, podtitl, disclaimer, veličina — sve podešljivo. Preview u real-time.
Sesije / razgovori	Pregled svih razgovora sa pretraživanjem, filterima po datumu/jeziku/intentu. Svaka poruka ima ocenu (thumb up/down) i metadata (IP, geo, vreme odgovora).
Bezbednost	Pregled sigurnosnih događaja, blokiranih poruka, rate limit događaja, detektovanih napada. Podešavanje pragova.
Tokeni	Generisanje i upravljanje klijentskim tokenima (JWT) za autentifikaciju widget-a.
Baza znanja	Pokretanje crawl-a, pregled KB statistike, status rebuild-a, pregled atoma znanja.
Korisnički nalozi	Više administratora sa 4 nivoa pristupa: viewer → operator → admin → superadmin. Svaki nivo ima tačno definisane dozvole za čitanje i pisanje.

5. Bezbednost

Funkcionalnost	Detalji
Prompt injection zaštita	SHA-256 kriptografski delimiteri oko sistemskog prompta. Napadač ne može da izvadi instrukcije niti da ih prepiše.
Detekcija napada	22+ pattern-a za detekciju pokušaja manipulacije AI-jem: prompt extraction, role override, delimiter attacks, encoding attacks.
Rate limiting	Konfigurabilan po IP-u i po sesiji. Burst zaštita (X poruka u Y sekundi) + ukupni dnevni limit. Automatski blokira abuse.
Input sanitizacija	Maksimalna dužina poruke: 2000 karaktera. Čišćenje opasnih patern-a pre nego što poruka dođe do AI-a.
Output filtriranje	AI odgovor prolazi kroz filter koji uklanja svaki pokušaj leak-a sistemskog prompta, internih tagova ili konfiguracije.
JWT autentifikacija	Svaki widget poziv zahteva validan JWT token. Tokeni se generišu sa SHA-256 derivacijom i ističu posle 24h.
Audit log	Svaka admin akcija se loguje: ko, šta, kada. Retencija: 180 dana.
Security log	Poseban log fajl za sigurnosne događaje. Retencija: 90 dana.

6. Chat widget

Widget je JavaScript fajl koji se ugrađuje jednom linijom koda na sajt klijenta. Radi u Shadow DOM-u — potpuno izolovan od CSS-a i skriptova sajta. Ne utiče na brzinu sajta.

Funkcionalnost	Detalji
Jedna linija koda	Ugradnja: <script src="..." data-endpoint="..." data-token="..."></script>. Radi na bilo kom sajtu, CMS-u ili platformi.
Trajna istorija	Razgovori se čuvaju u localStorage browsera. Korisnik može da zatvori tab, ugasi računar, vrati se sutradan — ceo razgovor je tu. Istorija se briše samo ako korisnik obriše cookie/localStorage.
Session ID	Generiše se jedinstven session ID pri prvom kontaktu i čuva se u localStorage. Svaki povratak istog korisnika nastavlja istu sesiju na serveru — AI pamti kontekst.
Typewriter efekat	AI odgovori se prikazuju karakter po karakter sa animiranim kursorom. Adaptivna brzina: kraći odgovori sporije (20ms), duži brže (10ms). Korisnik može da pošalje novu poruku i typewriter se odmah završi.
Typing indikator	Dok AI priprema odgovor, prikazuje se animirani "AI piše..." indikator sa tri pulsirajuće tačke.
Ocenjivanje (thumbs)	Svaki AI odgovor ima 👍 / 👎 dugmiće. Klik šalje ocenu na server. Vizuelni feedback: zeleno za dobro, crveno za loše. Jednom ocenjeno — ne može se promeniti.
Expand / Collapse	Na desktopu: dugme za proširenje panela na veću veličinu (80vw × 85vh). Na mobilnom: widget zauzima ceo ekran automatski.
Mobile fullscreen	Na ekranima ≤ 640px widget pokriva ceo ekran. Body scroll se zaključava (sprečava Safari bounce). iOS keyboard handling: koristi visualViewport API da pravilno prilagodi veličinu panela kad se pojavi tastatura.
Dark / Light tema i Formatiranje poruka	Podešavanje direktno na admin dashboard.AI odgovori podržavaju: **bold**, *italic*, URL-ovi automatski postaju klikabilni linkovi (target=_blank). HTML se sanitizuje (XSS zaštita).
Error handling	Mrežni greška: prikazuje lokalizovanu poruku (‘⚠️ Greška u vezi. Pokušajte ponovo.’). Input se ne zaključava — korisnik može odmah ponovo da pokuša.
Auto-open (opciono)	Podesiv data-auto-open atribut. Ako je uključen, widget se automatski otvara posle 2.8 sekundi — ali SAMO na desktopu, SAMO za nove korisnike koji nikad nisu zatvorili panel.

7. Integracije i deployment

Funkcionalnost	Detalji
Widget (JS)	Jedna linija koda za ugradnju na bilo koji sajt. Radi na svim browser-ima. Responsive dizajn. Typewriter efekat za odgovore. Thumb up/down za ocenjivanje.
WordPress plugin	Gotov .php plugin sa admin settings stranicom. Unese se endpoint URL + token i radi.
Email integracija	SMTP konfiguracija za slanje sales lead i support eskalacija emailova. Podržava bilo koji SMTP server.
Docker deployment	Multi-stage Dockerfile, docker-compose sa MongoDB i Redis. Auto-scaling Gunicorn workers (max 4). Health check endpoint. Max-requests za prevenciju memory leak-a.
MongoDB	Sva podešavanja, sesije, istorija razgovora, baza znanja, audit log — sve u MongoDB-u. Connection pooling (50 konekcija). TTL indeksi za automatsko čišćenje starih podataka.
Weather API	Open-Meteo integracija za vremenske komentare u pozdravu/oproštaju. Keširano 30 minuta. Geolocija korisnika iz IP adrese.

8. Pametne funkcije

Funkcionalnost	Detalji
Memorija (učenje)	Sistem pamti dobro ocenjene i loše ocenjene odgovore. Dobre koristi kao stilski uzor, loše izbegava. Memorija se gradi automatski od korisničkih i admin ocena. Admin ocene imaju veću težinu.
Vreme + pozdrav	AI generiše jedinstvene pozdrave i oproštaje bazirane na dobu dana i vremenskim uslovima. Nikad dvaput isti pozdrav. Fallback na template ako AI ne odgovori za 4 sekunde.
Sprečavanje halucinacija	6 slojeva zaštite: (1) zabrana izmišljanja cena, (2) zabrana izmišljanja kontakt podataka, (3) zabrana izmišljanja ličnih podataka korisnika, (4) regex verifikacija email/telefon, (5) code-level override LLM izlaza, (6) cenovnik verifikacija.
Tier-0 pravila	Obavezna pravila koja ne mogu biti prepisana od strane crawlovanog sadržaja. Definišu se po klijentu. Primeri: tačne SSH/SFTP konfiguracije, Windows hosting pravila, IP deblokada procedura.
Kontekst budget	Regularan KB sadržaj ograničen na 20.000 karaktera po upitu. Tier-0 pravila se nikad ne seču. Istorija razgovora: poslednjih 10 tura (20 poruka) - podesivo.
Klijentska konfiguracija	Svaki klijent ima sopstvenu konfiguraciju: persona, pravila, crawl URL-ovi, ključne reči za prodaju/podršku, obrasci simptoma, CTA tekstovi. Sve se menja iz admin panela bez ponovnog pokretanja.

9. Šta očekivati

Sistem radi:

- Odgovara na prodajna pitanja sa tačnim informacijama sa vašeg sajta
- Vodi korisnika kroz proces narudžbine do potvrde i email-a prodajnom timu
- Odgovara na tehnička pitanja iz baze znanja
- Uput ka živoj podršci kad ne može da pomogne
- Radi 24/7 bez pauze, na srpskom i engleskom
- Pamti šta dobro radi i unapređuje se
- Štiti se od zlonamernih korisnika i pokušaja manipulacije

Sistem ne radi:

- Ne zamenjuje kompleksnu tehničku podršku (L2/L3) — samo first-line
- Ne procesira plaćanja niti aktivira usluge — samo prikuplja podatke i prosledjuje
- Ne garantuje 100% tačnost AI odgovora — ali ima 6 slojeva zaštite od halucinacija
- Ne radi offline — zahteva internet konekciju, MongoDB i LLM API
- Kvalitet odgovora direktno zavisi od kvaliteta baze znanja (vaš sajt = AI-jevo znanje)

10. Tehničke specifikacije

Funkcionalnost	Detalji
Platforma	Python 3.11 + Flask + Gunicorn + gevent
Baza podataka	MongoDB 4.4+
Cache / queue	Redis
LLM	Bilo koji OpenAI-kompatibilan API (Ollama, OpenRouter, direktan API)
Deployment	Docker (docker-compose) ili bare metal
Web search	Brave Search API (opcionalno)
Email	SMTP (bilo koji provajder)
Weather	Open-Meteo API (besplatno, bez ključa)
Geolocija	IP-based geolocation (ip-api.com)
Logging	Loguru — stdout + dnevni fajlovi + sigurnosni log
Veličina koda	~45.000 linija (core: 15.000, knowledge engine: 15.000, templates + config: 15.000)

LIMARK DOO — Pančevo, Srbija — limark.rs — office@limark.rs — +381 60 139 6183