

FUNKCIONALNA SPECIFIKACIJA PROJEKTA

AI Baza znanja SKGO

Model isporuke:	client-hosted aplikacija
AI:	eksterni AI API kroz zamenjiv LLM Gateway
Hosting:	obezbeđuje SKGO
Fokus projekta:	kontrolisana AI baza znanja sa odgovorima iz validiranih dokumenata

1. Naziv projekta

AI Baza znanja SKGO

2. Svrha sistema

Cilj sistema je da zaposlenima u lokalnoj administraciji i SKGO omogući da kroz jednostavan chat interfejs postave pitanje iz oblasti pokrivena internom bazom dokumenata i dobiju odgovor zasnovan na važećim dokumentima koji su prethodno uneti u bazu samog sistema.

Ta dokumenta pokrivaju oblasti kao što su nacionalni i lokalni propisi, priručnici, modeli procedura i akata, analize, publikacije, instrukcije, i dr. Dokumenti se unose u obliku Doc, Excel, PDF i PPT fajlova.

Poseban set dokumenata čine validirani referentni materijali, uključujući primere pitanja i obrazloženih odgovora, koje sistem koristi kao deo baze znanja i konteksta za generisanje odgovora.“

Sistem treba da pomogne korisniku da:

- brzo pronađe relevantne propise, pravilnike, procedure, obrasce i uputstva;
- dobije razumljiv odgovor;
- vidi izvore na osnovu kojih je odgovor generisan;
- prepozna kada sistem nema dovoljno pouzdane informacije;
- dobije analizu, generisanu od sistema, o dokumentu koji je korisnik uneo i da mu sistem sugeriše njegovo unapređenje
- smanji potrebu za ručnim pretraživanjem dokumenata.

Sistem treba da razume pitanje, pretraži bazu znanja, pronađe relevantne dokumente, da odgovor, izvuče tačne delove i prikaže izvore kao što su naziv dokumenta, član, stranica i verzija dokumenta.

Modularnost i prilagodljivost

Sistem bi trebao biti razvijen tako da je moguće iskoristiti glavne funkcionalnosti u neke specijalizovane svrhe.

Ideja je da se i u budućnosti sistem koristi tako da se lako može dodati nova funkcija koja je specijalizovana upotreba osnovnog sistema.

3. Osnovni princip rada sistema

Sistem ne treba da funkcioniše kao slobodni chatbot koji odgovara iz opšteg znanja LLM modela koji koristi.

Sistem treba da radi po principu:

Pitanje korisnika

- pretraga validirane baze dokumenata
- izbor relevantnih segmenata
- AI odgovor zasnovan na izvorima
- prikaz izvora korisniku

Najvažnije pravilo:

Sistem ne sme da daje sadržinski odgovor ako nema dovoljno relevantnih izvora u bazi znanja.

U tom slučaju treba da prikaže poruku tipa:

“Nemam dovoljno pouzdanih informacija u bazi znanja za ovaj odgovor. Predlažem da uputite pitanje SKGO.”

4. Obim projekta

4.1. Funkcionalne celine

Projekat AI Baza znanja uključuje sledeće funkcionalne celine:

- korisnički web interfejs za postavljanje pitanja i AI chat sa odgovorima na osnovu baze znanja
- analiza korisnički unetih dokumenata;
- prikaz izvora odgovora;
- osnovna pretraga dokumenata;
- osnovni sistem za unos i obradu dokumenata — SUOD;
- metapodaci dokumenata;
- statusi dokumenata;
- osnovno indeksiranje dokumenata;
- RAG pipeline;
- LLM Gateway za povezivanje sa AI API providerom;
- admin panel;
- korisnici i osnovne role;
- token limiti i evidencija potrošnje;
- feedback korisnika;
- osnovna analitika;
- osnovni audit log;
- deployment na infrastrukturi klijenta.

4.2. Spremnost za nadogradnju

Projekat AI Baza znanja treba da bude izveden tako da omogući sledeće nadogradnje u narednim fazama:

- OCR slika;
- obradu skeniranih dokumenata bez tekstualnog sloja;
- potpun workflow odobravanja dokumenata;
- automatsko praćenje izmena zakona;
- opštu internet proveru odgovora;
- generisanje finalnih pravnih akata;
- integraciju sa ePisarnicom;
- integraciju sa DMS/SharePoint sistemom;
- napredni BI dashboard;
- višestepenu organizacionu hijerarhiju;

- potpuno lokalni AI model;
- enterprise high availability / disaster recovery;
- napredno upravljanje poverljivošću po dokumentu i korisniku.

5. Korisničke role

Potrebno je urediti tri osnovne role.

5.1. Krajnji korisnik

Krajnji korisnik je službenik koji koristi sistem za postavljanje pitanja.

Krajnji korisnik može da:

- pristupa samo frontu sistema, ne može pristupiti Sistemu za unos i obradu podataka (SUOD) kao ni Administrativnom panelu (AP)
- postavi pitanje;
- nastavi razgovor u okviru iste sesije;
- vidi odgovor;
- vidi izvore;
- otvori relevantne izvore;
- pretraži bazu dokumenata;
- oceni odgovor;
- prijavi problematičan odgovor;
- vidi svoju osnovnu istoriju razgovora.
- može da dodaje dokumente da bi zahtevao analizu od sistema

Krajnji korisnik ne može da:

- pristupi admin panelu
- pristupi SUOD-u
- menja dokumente;
- menja korisnike;
- menja systemske parametre;
- vidi sve korisničke upite drugih korisnika.

5.2. Administrator sadržaja

Administrator sadržaja upravlja dokumentima i bazom znanja.

Administrator sadržaja može da:

- uploaduje dokument;
- unese metapodatke;
- doda tagove;
- promeni status dokumenta;
- pokrene obradu/indeksiranje;
- deaktivira dokument;
- pregleda listu dokumenata;
- vidi osnovne greške pri obradi;
- vidi prijavljene odgovore i pitanja koja su ostala bez odgovora.

5.3. Sistemski administrator

Sistemski administrator upravlja korisnicima i osnovnim parametrima.

Sistemski administrator može da:

- kreira korisnike;

- deaktivira korisnike;
- dodeljuje role;
- podešava tipove korisnika;
- podešava token limite;
- pregleda potrošnju;
- podešava osnovne liste: oblasti, tipove dokumenata, tagove;
- vidi osnovne audit zapise.

6. Funkcionalna celina 1 — Korisnički chat

6.1. Opis

Korisnički chat je glavni interfejs za krajnje korisnike.

Korisnik unosi pitanje prirodnim jezikom. Sistem pretražuje bazu znanja, pronalazi relevantne segmente i generiše odgovor sa izvorima.

6.2. Funkcionalnosti

6.2.1. — Pokretanje novog razgovora

Korisnik može pokrenuti novi razgovor.

Sistem kreira novu chat sesiju.

Svaka sesija sadrži:

- ID sesije;
- korisnika;
- vreme početka;
- listu poruka;
- korišćene izvore;
- potrošnju tokena;
- status odgovora.

6.2.2. — Postavljanje pitanja

Korisnik može postaviti pitanje u slobodnom tekstu.

Primeri:

- “Koji dokumenti su potrebni za pokretanje postupka?”
- “Ko je nadležan za donošenje ove odluke?”
- “Koji rok se primenjuje u ovom slučaju?”
- “Da li je dokument koji ti prilažem adekvatno urađen”

6.2.3. — Nastavak razgovora

Pošto će sistem koristiti usluge eksternog LLMa, koji se naplaćuje po utrošku tokena, potrebno je omogućiti kontrolu utroška tokena za svakog korisnika. Inicijalno predvideti da Korisnik može postaviti najviše tri dopunska pitanja u okviru istog “posla”, ali omogućiti logiku i interfejs da se ta ograničenja lako koriguju, u skladu sa realnim stanjem potrošnje.

Treba predvideti ograničenja:

- jedan posao = početno pitanje + do 3 dopisivanja;
- nakon toga korisniku se može ponuditi da započne novi posao.

6.2.4. — Generisanje odgovora

Sistem generiše odgovor na osnovu pronađenih relevantnih segmenata.

Odgovor treba da sadrži:

- kratak zaključak;
- objašnjenje;
- korake procedure, ako su primenljivi;
- napomene ili ograničenja;
- izvore.

Poželjna struktura odgovora:

```

Kratki odgovor:
...

Objasnenje:
...

Koraci:
1. ...
2. ...
3. ...

Izvori:
- Dokument, član/tačka/strana, verzija

```

6.2.5. — Prikaz izvora

Svaki odgovor mora imati izvore ako je sistem dao sadržinski odgovor.

Izvor treba da prikaže:

- naziv dokumenta;
- tip dokumenta;
- relevantni član/tačku/odeljak ako postoji;
- stranicu ako je dostupna;
- verziju;
- status dokumenta;
- datum važenja ako je dostupan.

6.2.6. — Odgovor kada nema dovoljno informacija

Ako sistem ne pronađe dovoljno relevantnih izvora, ne sme da izmišlja odgovor.

Umesto toga prikazuje:

“Nemam dovoljno pouzdanih informacija u bazi znanja za ovaj odgovor. Predlažem da uputite pitanje SKGO ”

U sistemu se može prikazati i dugme:

“Prijavi kao pitanje bez odgovora”

ili

“Prosledi administratoru sadržaja”

6.2.7. — Status odgovora

Svaki AI odgovor dobija jedan od sledećih statusa:

Status	Značenje
Pouzdan odgovor	pronađeni su jasni izvori
Delimičan odgovor	postoje izvori, ali nisu dovoljni za konačan zaključak

Status	Značenje
Nema dovoljno informacija	sistem ne daje sadržinski odgovor

6.3. Analiza korisnički unetih dokumenata

Krajnji korisnik može da dostavi dokument radi analize njegovog sadržaja i dobijanja predloga za unapređenje dokumenta na osnovu podataka i informacija iz baze znanja (poređenjem sa dokumentima iz baze podataka i proverom usklađenosti priloženog dokumenta sa zahtevima, pravilima i kriterijumima u dokumentima iz baze znanja).

Dokument koji dostavlja krajnji korisnik:

- koristi se isključivo radi izvršenja konkretno zahtevane analize;
- čuva se u privremenom prostoru za skladištenje, odvojeno od validirane baze znanja;
- ne indeksira se u okviru baze znanja;
- ne može se koristiti kao izvor za odgovore drugim korisnicima;
- ne može se trajno uključiti u bazu znanja bez posebnog postupka unosa, provere i aktivacije od strane administratora sadržaja;
- automatski se briše najkasnije 24 časa nakon završetka analize, odnosno ranije ako korisnik prekine postupak ili obriše dokument.

Rezultat analize može se čuvati u okviru istorije razgovora korisnika, u skladu sa pravilima retencije chat logova.

Pre dostavljanja dokumenta korisniku se prikazuje upozorenje da ne sme da učitava dokumente koji sadrže poverljive podatke, podatke o ličnosti čija obrada nije dozvoljena, poslovne tajne, pristupne podatke, lozinke, API ključeve ili druge podatke čije dostavljanje eksternom AI API servisu nije izričito odobreno.

Korisnik mora potvrditi da je upoznat sa navedenim pravilima pre pokretanja analize.

7. Funkcionalna celina 2 — Pretraga baze znanja

7.1. Opis

Pored chata, korisnik treba da ima osnovnu pretragu dokumenata.

Razlog: u ovakvim sistemima korisnik često želi da vidi originalni dokument, a ne samo AI odgovor.

7.2. Funkcionalnosti

7.2.1 — Tekstualna pretraga

Korisnik može pretraživati dokumente po ključnim rečima.

7.2.2 — Prikaz rezultata

Rezultati pretrage prikazuju:

- naziv dokumenta;
- tip dokumenta;
- oblast;
- status;
- verziju;
- datum važenja;
- kratak relevantan isečak.

7.2.3 — Pregled dokumenta

Korisniku se prikazuje oprema (meta tagovi) dokumenta i link za preuzimanje.

Dovoljno je prikazati:

- metapodatke;
- relevantne tekstualne segmente;
- link ili dugme za preuzimanje originalnog fajla, ako je to dozvoljeno.

8. Funkcionalna celina 3 — SUOD: sistem za unos i obradu dokumenata

8.1. Opis

SUOD je modul za unos, obradu, klasifikaciju i indeksiranje dokumenata. Ovo je ključni modul jer kvalitet AI odgovora direktno zavisi od kvaliteta baze dokumenata.

8.2. Podržani tipovi fajlova u AI sistemu

Podržavaju se:

- PDF sa tekstualnim slojem;
- DOCX;
- TXT; PPT, PPTX,
- osnovni XLS, XLSX ako je tehnički jednostavan i strukturisan.

Ne podržavaju se:

- JPG/PNG slike;
- skenirani PDF bez OCR-a;
- rukom pisani dokumenti;

Podržani tipovi i veličine fajlova:

Maksimalna dozvoljena veličina pojedinačnog fajla mora biti tehnički podesiva kroz konfiguraciju sistema.

Inicijalno se primenjuju sledeća ograničenja:

- dokument koji administrator sadržaja unosi u SUOD: najviše 50 MB po fajlu;
- dokument koji krajnji korisnik dostavlja radi analize: najviše 20 MB po fajlu.

Sistem mora odbiti upload fajla koji prelazi dozvoljenu veličinu i prikazati korisniku jasnu poruku o razlogu odbijanja.

8.3. Upload dokumenta

Administrator sadržaja može uploadovati dokument.

Pri uploadu sistem evidentira:

- naziv fajla;
- format;
- veličinu;
- korisnika koji je izvršio upload;
- datum i vreme uploada;
- inicijalni status obrade.

8.4. Metapodaci dokumenta

Za svaki dokument se unose sledeći metapodaci:

Polje	Obavezno	Opis
Naziv dokumenta	Da	službeni naziv
Tip dokumenta	Da	zakon, pravilnik, procedura, obrazac, FAQ itd.
Oblast	Da	oblast kojoj dokument pripada
Podoblast	Ne	uža klasifikacija
Tagovi	Da	oznake za pretragu
Verzija	Da	oznaka verzije
Status	Da	aktivan, neaktivan, arhiviran
Izvor	Da	SKGO, nacionalni organi, lokalni organi, drugi izvori
Datum donošenja	Ne	ako je poznat
Datum stupanja na snagu	Ne	ako postoji
Datum prestanka važenja	Ne	ako postoji
Vlasnik dokumenta	Ne	organizacija ili služba
Kratak opis	Ne	opis sadržaja
Napomena	Ne	dodatne informacije

8.5. Statusi dokumenata

U projektu AI Baze se uvode sledeći statusi:

Status	Značenje
U obradi	dokument je uploadovan i obrađuje se
Aktivan	dokument može biti korišćen kao izvor
Neaktivan	dokument postoji, ali se ne koristi u odgovorima
Arhiviran	dokument se čuva, ali nije važeći izvor

Sistem u AI odgovorima sme koristiti samo dokumente sa statusom: **Aktivan**

Potrebno je omogućiti da dokumenti sa statusom U OBRADI se mogu naknadno učitati i nastaviti rad sa njihovom obradom.

8.6. Obrada dokumenta

Nakon uploada sistem pokreće obradu:

- čitanje dokumenta;
- ekstrakcija teksta;
- osnovna validacija da li dokument ima tekst;
- segmentacija;
- opis metapodataka;
- kreiranje embeddinga;
- opis u vector/text search indeks;
- označavanje dokumenta kao spremnog za korišćenje.

8.7. Segmentacija dokumenta

Dokument se deli na manje celine.

Koristi praktična segmentacija:

- po naslovima ako su dostupni;
- po pasusima;
- po približnoj dužini teksta;

- uz čuvanje reference na dokument i poziciju segmenta.

Za kasniju fazu može se uvesti preciznije deljenje po članovima, tačkama i odeljcima.

8.8. Tagovi

Administrator može dodati tagove dokumentu.

Tagovi mogu biti:

- unapred definisani;
- ručno dodati;
- predloženi od strane AI sistema i po odobrenju administratora unosa dodati dokumentu.

Za projekat AI Baza znanja:

- ručni tagovi su uključeni;
- AI predlog tagova je opciona stavka ako ostane prostora u budžetu;
- automatsko kreiranje novih tagova bez potvrde čoveka nije uključeno.

8.9. Reindeksiranje dokumenta

Administrator može ručno pokrenuti reindeksiranje dokumenta.

Koristi se kada:

- promenjeni su metapodaci;
- promenjen je status;
- uploadovana je nova verzija;
- došlo je do greške u prethodnoj obradi.

9. Funkcionalna celina 4 — RAG / AI pipeline

9.1. Opis

RAG pipeline je centralni AI mehanizam sistema.

On omogućava da sistem odgovara na osnovu baze znanja, a ne na osnovu slobodnog znanja AI modela.

9.2. Tok obrade pitanja

Kada korisnik postavi pitanje, sistem radi sledeće:

- prihvata pitanje;
- proverava korisnička prava i token limite;
- priprema upit za pretragu;
- pretražuje vector/text indeks;
- filtrira samo aktivne dokumente;
- rangira relevantne segmente;
- proverava da li ima dovoljno izvora;
- šalje pitanje i relevantne segmente AI API-ju;
- dobija odgovor;
- validira da li odgovor ima izvore;
- prikazuje odgovor korisniku;
- evidentira potrošnju i izvore.

9.3. Pretraga izvora

Sistem koristi kombinaciju:

- semantičke/vector pretrage;

- tekstualne pretrage;
- metadata filtera.

Mora se implementirati osnovni mehanizam rangiranja po relevantnosti.

9.4. Pravila za izbor izvora

Sistem bira izvore prema:

- relevantnosti za pitanje;
- statusu dokumenta;
- oblasti;
- tipu dokumenta;
- verziji;
- datumu važenja;
- prioritetu dokumenta.

Potrebno je razraditi sistem prioriteta dokumenata, na primer da zakon ima veću važnost od interne procedure.

Za projekat AI Baza znanja se uvodi sledeća hijerarhija prioriteta:

Prioritet	Tip dokumenta
1	zakon
2	uredba
3	pravilnik
4	priručnici, vodiči i instrukcije
5	mišljenja relevantnih organa
6	modeli dokumenata
7	interna procedura
8	FAQ
9	materijali sa obuka i prezentacije
0	specijalni dokument

U slučaju kada sistem pronađe izvore čiji sadržaj nije međusobno usklađen, sistem najpre pokušava da razreši konflikt primenom sledećih kriterijuma:

- status dokumenta — koriste se samo aktivni dokumenti;
- datum važenja — prednost ima dokument koji je važeći u relevantnom periodu;
- verzija dokumenta — prednost ima novija važeća verzija;
- hijerarhija prioriteta izvora;
- pravna snaga dokumenta;
- datum donošenja ili izmene dokumenta.

Ako se konflikt može pouzdano razrešiti primenom navedenih kriterijuma, sistem daje odgovor na osnovu relevantnog izvora i, kada je to korisno, navodi da postoji dokument drugačijeg sadržaja koji nema prioritet u konkretnom slučaju.

Ako sistem ne može pouzdano da razreši konflikt, ne daje konačan sadržinski odgovor.

9.5. Pravila odgovaranja

AI model dobija instrukciju da:

- odgovara samo na osnovu dostavljenih izvora koji se nalaze u bazi znanja;
- ne izmišlja članove, rokove, procedure ili pravila;

- navodi kada nema dovoljno informacija;
- ne koristi opšte znanje ako nije potvrđeno izvorima;
- ne daje konačan pravni akt;
- koristi oprezan i stručan ton.

9.6. Fallback logika

Sistem ne generiše sadržinski odgovor ako:

- nema pronađenih relevantnih izvora;
- izvori su prenisko rangirani;
- svi izvori su neaktivni ili arhivirani;
- postoji očigledan konflikt izvora;
- pitanje je van obuhvata baze znanja;
- pitanje traži lične, poverljive ili neprimerene podatke.

Sistem ne generiše konačan sadržinski odgovor kada pronađeni aktivni izvori sadrže međusobno suprotstavljene informacije, a konflikt nije moguće pouzdano razrešiti primenom pravila za izbor izvora.

U tom slučaju sistem prikazuje poruku:

„U bazi znanja postoje izvori sa različitim informacijama, zbog čega nije moguće dati pouzdan konačan odgovor. Pitanje je evidentirano radi dodatne provere.“

Sistem prikazuje pronađene konfliktne izvore i automatski evidentira pitanje sa statusom „potrebna provera“

10. Funkcionalna celina 5 — LLM Gateway

10.1. Opis

LLM Gateway je interni sloj koji omogućava da aplikacija ne zavisi direktno od jednog AI providera.

10.2. Funkcionalnosti

LLM Gateway treba da podrži:

- konfiguraciju AI providera;
- konfiguraciju modela;
- slanje prompta;
- prijem odgovora;
- merenje token potrošnje;
- čuvanje modela korišćenog za odgovor;
- timeout;
- retry logiku;
- evidentiranje grešaka;
- mogućnost zamene providera bez promene glavne aplikacije.

10.3. Konfiguracija modela

Administrator sistema ili tehnički administrator može podesiti:

- naziv providera;
- API endpoint;
- API ključ;
- naziv modela;
- maksimalan broj input tokena;

- maksimalan broj output tokena;
- timeout;
- dnevni/mesečni budžetski limit.

Ova konfiguracija se može držati u tehničkoj konfiguraciji, ne mora imati kompletan UI.

11. Funkcionalna celina 6 — Admin panel

11.1. Opis

Admin panel omogućava upravljanje dokumentima, korisnicima, tokenima i osnovnim parametrima sistema.

Početni dokument predviđa da admin panel bude dostupan samo administratorima sistema i da se u njemu kreiraju korisnički nalozi i podešavaju parametri sistema.

11.2. Upravljanje korisnicima

Uvode se osnovni tipovi korisnika:

- standardni
- plus
- super

Administrator može:

- videti listu korisnika;
- kreirati korisnika;
- izmeniti korisnika;
- deaktivirati korisnika;
- dodeliti rolu;
- resetovati pristup ako se koristi lokalna autentifikacija.

11.3. Upravljanje token limitima

Administrator može podesiti:

- dnevni limit;
- mesečni limit;
- limit po tipu korisnika;
- limit po pojedinačnom korisniku.

Sistem prati:

- broj AI poziva;
- input tokene;
- output tokene;
- ukupnu procenjenu potrošnju;
- datum i vreme korišćenja.

11.4. Upravljanje listama

Administrator sadržaja može upravljati osnovnim šifarnicima:

- oblasti;
- podoblasti;
- tipovi dokumenata;
- tagovi;
- statusi dokumenata.

11.5. Pregled pitanja bez odgovora

Administrator sadržaja može videti pitanja za koja sistem nije dao odgovor.

Za svako pitanje prikazuje se:

- korisnik;
- vreme;
- tekst pitanja;
- status;
- razlog neodgovaranja;
- eventualno pronađeni slabi izvori;
- feedback korisnika.

12. Funkcionalna celina 7 — Analitika i izveštaji

12.1. Opis

Projekat AI Baza znanja uključuje osnovnu analitiku dovoljnu za kontrolu troškova.

12.2. Osnovni pokazatelji

Sistem prikazuje:

- broj korisnika;
- broj pitanja;
- broj AI odgovora;
- broj pitanja bez odgovora;
- broj negativnih feedback-a;
- token potrošnju po danu;
- token potrošnju po korisniku;
- najčešće korišćene dokumente;
- najčešće oblasti pitanja.

12.3. Izvoz podataka

Za ovu fazu projekta AI Baza znanja je potrebno omogućiti jednostavan CSV, PDF i DOCX export za:

- pitanja;
- token potrošnju;
- feedback;
- pitanja bez odgovora.
- rezultat analize korisnički unetog dokumenta

13. Funkcionalna celina 8 — Feedback i osnovna eskalacija

13.1. Feedback

Korisnik nakon odgovora može označiti:

- odgovor je koristan;
- odgovor nije koristan;
- izvor nije dobar;
- odgovor je nepotpun;

13.2. Pitanja bez odgovora

Ako sistem ne može da odgovori, pitanje se evidentira u posebnoj listi.

Administrator sadržaja može pregledati listu i odlučiti:

- da treba dodati dokument;
- da treba dopuniti metapodatke;
- da je pitanje van obuhvata;
- da pitanje treba proslediti domenskom ekspertu.

13.3. Osnovna eskalacija

Za projekata AI Baza znanja eskalacija ne mora biti pun ticketing sistem.

Dovoljno je:

- označiti pitanje kao “potrebna provera”;
- prikazati ga administratoru;
- omogućiti unos interne napomene;
- označiti status: otvoreno, u obradi, zatvoreno.

Pitanje se automatski označava statusom „potrebna provera“ i prikazuje administratoru sadržaja kada:

- sistem nema dovoljno pouzdane izvore;
- pronađeni izvori sadrže konflikt koji sistem ne može pouzdano da razreši;
- korisnik prijavi problematičan odgovor;
- pitanje zahteva dodatnu proveru domenskog eksperta.

Za pitanja sa konfliktom izvora sistem mora prikazati listu pronađenih izvora i razlog zbog kojeg konflikt nije automatski razrešen.

14. Funkcionalna celina 9 — Audit log

14.1. Opis

Sistem mora čuvati osnovni audit trag za ključne aktivnosti.

14.2. Aktivnosti koje se loguju

Loguju se:

- login korisnika;
- postavljeno pitanje;
- generisan odgovor;
- korišćeni izvori;
- AI model koji je korišćen;
- token potrošnja;
- upload dokumenta;
- promena statusa dokumenta;
- reindexiranje dokumenta;
- kreiranje korisnika;
- promena role;
- promena token limita;
- feedback korisnika.

14.3. Retencija logova

Omogućiti tehnički podesivu dužinu čuvanja logova..

Inicijalno predviđeno:

- chat logovi: 90–180 dana;
- audit logovi: 12 meseci;
- token logovi: 12 meseci.
- dokumenta koja na analizu prilaže korisnik: 24 sata

15. Nefunkcionalni zahtevi

15.1. Client-hosted deployment

Aplikacija se instalira na infrastrukturu koju obezbeđuje SKGO.

SKGO obezbeđuje:

- servere ili VM-ove;
- bazu podataka;
- storage za dokumente;
- backup;
- mrežni pristup;
- firewall pravila;
- monitoring infrastrukture;
- pristup AI API endpointu.

Projektni tim obezbeđuje:

- aplikativni kod;
- deployment uputstva;
- konfiguraciju aplikacije;
- Docker/deployment pakete ako se tako dogovori;
- pomoć pri instalaciji;
- integraciju sa AI API-jem.

15.2. Bezbednost

Zahtevaju se:

- HTTPS;
- autentifikacija;
- role-based access control;
- zaštita API ključeva;
- ograničenje pristupa admin panelu;
- audit log;
- osnovna zaštita od neautorizovanog pristupa;
- validacija fajlova pri uploadu;
- ograničenje veličine fajlova;
- zabrana obrade nepodržanih fajlova.

15.3. Performanse

Projekat AI Baza znanja cilja:

- odgovor na jednostavno pitanje: do 10–20 sekundi;
- odgovor na složenije pitanje: do 30 sekundi;
- upload dokumenta: zavisi od veličine dokumenta;
- obrada dokumenta može biti asinhrona.

Ako AI API ili lokalna infrastruktura klijenta kasne, sistem mora prikazati korisniku jasnu poruku.

15.4. Skalabilnost

Projekat treba projektovati tako da kasnije može podržati:

- veći broj korisnika;
- više dokumenata;
- više oblasti;
- naprednije role;
- više AI providera;
- eventualni lokalni AI model.

Projekat treba da podrži:

- Ukupan broj korisnika:
- Broj istovremenih korisnika:
- Broj dnevnih korisnika:

15.5. Jezik i pismo

Sistem mora podržati srpski jezik i rad sa ćiriličnim i latiničnim pismom. Za potrebe pretrage i pronalaženja izvora, sistem mora obezbediti normalizaciju teksta tako da korisnički upit može biti unet jednim pismom, a relevantni dokument pronađen i kada je napisan drugim pismom. Interfejs sistema, osnovne sistemske poruke i obaveštenja isporučuju se na srpskom jeziku.

15.6. Vizuelni identitet donatora projekta

U vizuelnom identitetu sistema mora biti naznačeno da je proizvod nastao u okviru projekta "Partnerstvo za dobru lokalnu samoupravu" ,koji sprovodi SKGO uz podršku Vlade Švajcarske, Oznaka se prikazuje na prijavnom ekranu i u podnožju (footer) aplikacije, u skladu sa pravilima vidljivosti donatora.

16. Ograničenja Projekta AI Baza znanja

Element	Ograničenje
Broj oblasti	biće kasnije definisano
Broj dokumenata	20 000
Broj ukupnih korisnika	10 000
Broj dnevnih korisnika	1000
Broj istovremenih korisnika	500
Tipovi fajlova	PDF tekstualni, DOCX, TXT, XLS, PPT, PPTX, XLSX,
AI provider	jedan primarni + mogućnost zamene kroz gateway

17. Kriterijumi prihvatanja

Projekat AI Baza znanja se smatra uspešno isporučenim ako su ispunjeni sledeći uslovi.

17.1. Korisnički chat

- korisnik može da se prijavi;
- korisnik može da pokrene razgovor;
- korisnik može da postavi pitanje;
- sistem vraća odgovor;
- odgovor prikazuje izvore;
- korisnik može da nastavi razgovor;

- korisnik može dati feedback.
- Korisnik može da postavi dokument i zatraži analizu dokumenta
- Sistem vraća odgovor i traženu analizu

17.2. Izvori

- svaki sadržinski odgovor ima barem jedan izvor;
- izvor sadrži naziv dokumenta;
- ako je dostupno, prikazuje se član/odeljak/strana;
- sistem ne koristi neaktivne dokumente kao izvor;
- sistem jasno kaže kada nema dovoljno informacija.

17.3. Dokumenti

- admin sadržaja može uploadovati dokument;
- admin sadržaja može uneti metapodatke;
- dokument se može obraditi;
- dokument se može indeksirati;
- dokument može dobiti status aktivan/neaktivan/arhiviran;
- neaktivan/arhiviran dokument se ne koristi u odgovorima.

17.4. Sistemski Administrator i Administrator sadržaja

- SA može upravljati korisnicima;
- AS može upravljati dokumentima;
- AS može upravljati osnovnim tagovima i oblastima;
- SA može videti token potrošnju;
- AS može videti pitanja bez odgovora.

17.5. AI/RAG

- sistem pronalazi relevantne segmente;
- sistem šalje samo relevantne segmente AI modelu;
- sistem vraća odgovor zasnovan na izvorima;
- sistem ne odgovara kada nema dovoljno izvora;
- potrošnja tokena se evidentira.

17.6. Testiranje

- postoji test set od najmanje 50 pitanja;
- za test pitanja su označeni očekivani izvori;
- testiraju se i pitanja na koja sistem ne sme da odgovori;
- domain expert pregleda uzorak odgovora.

19. Planirana struktura ekrana

19.1. Korisnički deo

Ekрани:

- Login (username i password)
- Početna strana
- Novi chat
- Istorija razgovora
- Prikaz odgovora sa izvorima

- Pretraga dokumenata
- Detalj dokumenta
- Feedback / prijava problema

19.2. Admin deo – Sistemski administrator i Administrator sadržaja

SA Ekрани:

- Admin dashboard
- Lista dokumenata
- Upload dokumenta
- Forma za metapodatke
- Detalj dokumenta
- Lista korisnika
- Kreiranje/izmena korisnika
- Token potrošnja
- Pitanja bez odgovora
- Feedback lista
- Osnovna podešavanja

AS Ekрани:

- Lista dokumenata
- Upload dokumenta
- Forma za metapodatke
- Detalj dokumenta
- Pitanja bez odgovora
- Feedback lista

20. Tokovi rada

20.1. Tok: korisnik postavlja pitanje

- Korisnik se prijavi.
- Otvara novi chat.
- Postavlja pitanje.
- Sistem proverava limit.
- Sistem pretražuje bazu.
- Sistem pronalazi izvore.
- Sistem šalje relevantne segmente AI modelu.
- Sistem vraća odgovor.
- Korisnik vidi odgovor i izvore.
- Korisnik daje feedback.

20.2. Tok: nema dovoljno informacija

- Korisnik postavlja pitanje.
- Sistem pretražuje bazu.
- Relevantnost izvora je niska ili nema aktivnih izvora.
- Sistem ne generiše sadržinski odgovor.
- Prikazuje poruku da nema dovoljno informacija.
- Pitanje se evidentira kao “bez odgovora”.
- Administrator ga vidi u admin panelu.

20.3. Tok: unos dokumenta

- Administrator uploaduje dokument.
- Unosi metapodatke.
- Sistem proverava format.
- Sistem ekstrahuje tekst.
- Sistem segmentira dokument.
- Sistem indeksira dokument.
- Administrator postavlja status “aktivan”.
- Dokument postaje dostupan za AI odgovore.

20.4. Tok: deaktivacija dokumenta

- Administrator otvara dokument.
- Menja status u “neaktivan” ili “arhiviran”.
- Sistem reindeksira ili izuzima dokument iz aktivnog indeksa.
- Dokument više nije dostupan kao izvor u AI odgovorima.

20.5. Tok: analiza dokumenta

- Korisnik uploaduje dokument.
- Korisnik zatraži analizu dokumenta.
- Sistem proverava format.
- Sistem proverava limit.
- Sistem pretražuje bazu.
- Sistem pronalazi izvore.
- Sistem šalje relevantne segmente AI modelu.
- Sistem vraća odgovor.
- Korisnik vidi odgovor i izvore.
- Korisnik daje feedback.

20.6. Tok: pretraga baze znanja iz chat polja

- Korisnik u chat polje upisuje zahtev da sistem pretraži bazu znanja i pronađe određeni dokument, propis, proceduru, obrazac, priručnik ili drugi relevantan materijal.
- Sistem prepoznaje da korisnik ne traži generisanje sadržinskog odgovora, već pretragu i pronalaženje dokumenata iz validirane baze znanja.
- Sistem proverava korisnička prava, status korisnika i dostupnost funkcionalnosti pretrage baze znanja.
- Sistem priprema upit za pretragu, uključujući normalizaciju ćirilicnog i latiničnog pisma, izdvajanje ključnih reči i primenu dostupnih metadata filtera.
- Sistem pretražuje bazu znanja kombinacijom tekstualne pretrage, semantičke odnosno vektorske pretrage i metadata filtera, ako su primenljivi.
- Sistem u rezultatima uzima u obzir samo dokumente koji su dostupni krajnjem korisniku i koji imaju status aktivan.
- Sistem rangira pronađene dokumente prema relevantnosti, tipu dokumenta, oblasti, verziji, datumu važenja i prioritetu dokumenta.
- Ako sistem pronađe jedan jasno najrelevantniji dokument, korisniku prikazuje taj dokument kao preporučeni rezultat.
- Ako sistem pronađe više relevantnih dokumenata, korisniku prikazuje listu rezultata sortiranu po relevantnosti.
- Za svaki rezultat sistem prikazuje najmanje naziv dokumenta, tip dokumenta, oblast, status, verziju, datum važenja ako postoji
- Korisniku se omogućava preuzimanje originalnog fajla ako je to dozvoljeno pravilima sistema.

21. Podaci koji se čuvaju

21.1. Korisnik

- ID;
- ime i prezime;
- email;
- rola (standard, plus, super);
- status;
- token limit;
- datum kreiranja;
- poslednja aktivnost.

21.2. Dokument

- ID;
- naziv;
- tip;
- oblast;
- tagovi;
- verzija;
- status;
- datum važenja;
- vlasnik;
- originalni fajl;
- ekstrahovani tekst;
- segmenti;
- datum obrade;
- greške obrade.

21.3. Chat sesija

- ID;
- korisnik;
- vreme početka;
- vreme poslednje aktivnosti;
- status;
- potrošnja tokena;
- broj poruka.

21.4. AI odgovor

- pitanje;
- odgovor;
- izvori;
- model;
- input tokeni;
- output tokeni;
- vreme odgovora;
- status odgovora;

- feedback.

22. Prioriteti za razvoj

Prioritet 1 — Mora biti urađeno

- login;
- korisnički chat;
- RAG odgovor;
- izvori;
- fallback bez izvora;
- upload dokumenata;
- analiza dokumenata
- metapodaci;
- status dokumenta;
- indeksiranje;
- admin korisnici;
- token evidencija;
- deployment kod klijenta;
- osnovno testiranje.

Prioritet 2 — Važno, ali može biti pojednostavljeno

- klasična pretraga;
- istorija razgovora;
- feedback;
- pitanja bez odgovora;
- osnovna analitika;
- ručni tagovi;
- CSV export.

23. Tehnička struktura projekta

Sloj	Osnovni plan
Frontend	Moderna web aplikacija, single-page ili server-assisted frontend
Backend	Modularni poslovni backend / modularni monolit
Baza	Relaciona baza sa podrškom za strukturirane podatke i pretragu
Vector search	U prvoj fazi integrisan vector search uz bazu; kasnije moguć poseban vector/search servis
Dokument storage	Zamenjiv storage sloj: lokalni, object storage ili cloud storage
AI sloj	LLM Gateway sa podrškom za više AI providera

Embeddings	Zamenjiv embedding provider: eksterni API ili lokalni model
Auth	Interni auth za MVP, SSO-ready arhitektura
Background obrada	Asinhroni worker/job sistem za obradu dokumenata
Deployment	Container-ready deployment
Monitoring	Osnovni health check, logovi, audit i usage metering
Security	RBAC, audit, enkripcija u transportu, zaštita API ključeva
Tenant model	Single-tenant ali tenant-ready data model

24. Intelektualna prava i operativna nezavisnost SKGO

SKGO je vlasnik izvornog koda, strukture baze znanja, klasifikacije i strukture podataka, matapodataka, upita (promptova) i drugih tehničkih dokumenata (analize, izveštaji, izvedeni zaključci i dr.) koji su posebno pripremljeni ili obrađeni u procesu uspostavljanja sistema „AI Baza znanja SKGO“. Izvođač mora da obezbedi da SKGO nakon završetka ugovora može da koristi sistem, menja konfiguraciju i angažuje drugog izvođača za održavanje i dalji razvoj.

Nije prihvatljivo tehničko rešenje koje bez opravdanog razloga stvara trajnu zavisnost SKGO od jednog izvođača, AI provajdera ili vlasničke komponente treće strane.

Izvođač predaje SKGO:

1. kompletan izvorni kod aplikacije, uključujući sve module i komponente razvijene u okviru projekta;
2. migracije baze podataka;
3. deployment skripte, konfiguracione šablone i container-ready pakete;
4. tehničku dokumentaciju sistema, koja obuhvata:
 - opis arhitekture sistema i ključnih komponenti;
 - opis toka podataka (dataflow) kroz ključne komponente sistema od unosa dokumenta, preko ekstrakcije, segmentacije i indeksiranja, do generisanja odgovora;
 - detaljan opis procesa obrade, indeksiranja i ponovnog indeksiranja dokumenata;
 - dokumentaciju baze podataka;
 - dokumentaciju javnih i internih API-ja;
 - dokumentaciju programskih celina (moduli/namespaces, tipovi, metode, funkcije i konstante), na nivou dovoljnom za samostalno održavanje i dalji razvoj sistema od strane drugog izvođača;
5. spisak svih komponenti trećih lica, sa pregledom pripadajućih licenci i uslova korišćenja;
6. operativnu dokumentaciju, koja obuhvata:
 - uputstvo za instalaciju i konfiguraciju sistema;
 - uputstvo za backup i restore;
 - uputstvo za monitoring sistema;
 - uputstvo za ponovno indeksiranje dokumenata;
 - uputstvo za promenu AI provajdera i embeddings rešenja;
7. administratorsko uputstvo;
8. korisničko uputstvo.

SKGO će na izvršioca preneti isključivo pravo da opšta znanja, iskustvo, koncepte, arhitektonske obrasce, algoritamske pristupe, generičke komponente, biblioteke, šablone, izvorni kod i druga tehnička rešenja, koja su razvijena okviru ovog projekta koristi i dalje razvija za sopstvene proizvode uz ograničenjhe da proizvode zasnovane tome prodaje/ustupa/stavi na raspolaganje državnim organima, organima autonomne pokrajine i organima lokalne samouprave, odnosno institucijama/organizacijama osnovanim od strane istih, bez pisane saglasnosti SKGO.

ANEKS 1

FAZE I ROKOVI REALIZACIJE I PLAĆANJA USLUGA I KRITERIJUMI I NAČIN PRIHVATANJA OSTVARENJA PLANIRANIH REZULTATA

AI Baza znanja SKGO

1. Predmet i svrha priloga

Ovim prilogom uređuju se faze realizacije projekta „AI Baza znanja SKGO“ odnosno proizvodi koje je izvođač dužan da isporuči u svakoj fazi, način provere kvaliteta, kriterijumi prihvatanja i dinamika plaćanja.

Prilog predstavlja sastavni deo projektne i ugovorne dokumentacije i primenjuje se zajedno sa Funkcionalnom specifikacijom projekta „AI Baza znanja SKGO“. U slučaju potrebe za tumačenjem obima pojedinačne funkcionalnosti, polazi se od Funkcionalne specifikacije, ovog priloga i pisano prihvaćene detaljne tehničke specifikacije iz Faze 1 ovog Aneksa.

2. Osnovna načela realizacije

- Realizacija se sprovodi po fazama, sa jasnom dinamikom isporukai kriterijumima prihvatanja.
- Isplata se vrši periodično, na osnovu formalno prihvaćene odgovarajuće faze
- Sistem se razvija kao client-hosted aplikacija i instalira na infrastrukturi koju obezbeđuje SKGO.
- Sistem mora biti modularan, dokumentovan i nezavisan od jednog AI provajdera kroz zamenjivi LLM Gateway.
- Promene obima nakon prihvatanja detaljne specifikacije sprovode se isključivo kroz pisano odobren postupak upravljanja promenama.

3. Pregled faza i dinamika plaćanja

Faza	Ključni rezultat	Udeo u ugovorenoj ceni
1. Usvojena funkcionalna i tehnička specifikacija	Formalno potvrđena specifikacija i arhitektura	15%
2. Završena Core platforma i osnovna infrastruktura	Backend osnova, autentifikacija, role, dokument menadžment, deployment osnova	25%
3. Završena AI/RAG funkcionalnost	Upload, indeksiranje, pretraga, AI odgovori, citiranje izvora, osnovna evaluacija	25%
4. Završena aplikacija za korisničko prihvatno testiranje (UAT)	Dostavljena kompletna verzija sa: Korisnički chat, admin panel, upravljanje dokumentima, prava pristupa Aplikacija spremna za korisničko testiranje	15%
5. Uspešno završen UAT	Ispravljene kritične greške i prihvaćena testna verzija. QA, bug fixing, sigurnosne provere, performanse, priprema produkcije	10%

Faza	Ključni rezultat	Udeo u ugovorenoj ceni
6. Go-live i finalna primopredaja	Produkciono puštanje, dokumentacija, obuka i finalna primopredaja	10%
UKUPNO		100%

4. Faze realizacije i kriterijumi prihvatanja

4.1. Faza 1 – Usvojena funkcionalna i tehnička specifikacija

Cilj faze je da se pre početka razvoja nedvosmisleno utvrde obim, arhitektura, tehničke pretpostavke, režim zaštite podataka i način provere rezultata.

Kriterijum prihvatanja

Faza se prihvata kada SKGO pisanim putem odobri detaljnu tehničku specifikaciju, matricu obuhvata i plan testiranja. Odobrena dokumentacija predstavlja osnov za realizaciju narednih faza i kontrolu eventualnih promena.

4.2. Faza 2 – Završena Core platforma i osnovna infrastruktura

Cilj faze je uspostavljanje funkcionalne osnove aplikacije i tehničkog sloja koji omogućava kontrolisano povezivanje sa eksternim AI API servisom (backend osnova, autentifikacija, osnovne role i prava pristupa, početni admin panel, osnovni dokument menadžment, osnovna evidencija token potrošnje, LLM Gateway sa demonstriranom konfigurabilnošću AI providera/modela i deployment osnova).

Kriterijum prihvatanja

Izvođač u testnom okruženju demonstrira rad korisničkog naloga i uspešan poziv najmanje jednog eksternog AI API-ja preko LLM Gateway-a. Zamena modela ili provajdera mora biti moguća kroz konfiguraciju, bez promene osnovne poslovne logike aplikacije.

4.3. Faza 3 – AI / RAG funkcionalnost

Ova faza predstavlja centralnu funkcionalnu celinu projekta. U ovoj fazi se očekuje - upload, obrada, segmentacija, indeksiranje i reindexiranje dokumenata; pretraga aktivne baze znanja; AI/RAG odgovori zasnovani na validiranim izvorima; citiranje izvora; fallback logika; osnovna evaluacija; analiza korisnički unetog dokumenta u privremenom režimu, bez njegovog uključivanja u validiranu bazu znanja.

U okviru ove faze proverava se i funkcionalnost sistema za unos i obradu dokumenata u meri potrebnoj za rad AI/RAG funkcionalnosti. To obuhvata upload podržanih tipova fajlova, unos i čuvanje metapodataka, dodelu statusa dokumenta, ekstrakciju teksta, segmentaciju, kreiranje embeddinga, indeksiranje, filtriranje samo aktivnih dokumenata i ručno reindexiranje dokumenta.

U okviru ove faze proverava se i funkcionalnost analize dokumenta koji unosi krajnji korisnik. Sistem mora omogućiti da krajnji korisnik učitava dokument radi konkretne analize, da se dokument koristi isključivo za tu analizu, da se ne indeksira u validiranu bazu znanja, da se ne koristi kao izvor za odgovore drugim korisnicima i da sistem korisniku vrati analizu dokumenta sa pozivanjem na relevantne izvore iz validirane baze znanja

Sistem mora da generiše odgovor isključivo na osnovu pronađenih i validiranih izvora iz baze znanja.

Obavezna fallback logika

Sistem ne generiše sadržinski odgovor kada nema relevantnih izvora, kada su izvori prenisko rangirani, kada su pronađeni samo neaktivni ili arhivirani dokumenti, kada postoji očigledan konflikt izvora, kada je pitanje van obuhvata baze ili kada pitanje traži lične, poverljive ili neprimerene podatke.

U navedenim slučajevima sistem prikazuje jasnu poruku da nema dovoljno pouzdanih informacija u bazi znanja i omogućava evidentiranje pitanja radi dalje provere.

Kriterijum prihvatanja

Faza se prihvata nakon provere testnog seta koji posebno pitanja sa jasnim odgovorom, pitanja sa više relevantnih izvora, pitanja sa konfliktnim izvorima, pitanja iz nepokrivenih oblasti, pitanja za koja postoje samo neaktivni dokumenti i pokušaje dobijanja odgovora iz opšteg znanja AI modela, i analiza korisnički unetog dokumenta.

Faza se prihvata ako sistem uspešno demonstrira celokupan tok od unosa dokumenta u sistem za unos i obradu podataka (SUOD), preko obrade, segmentacije, indeksiranja i aktivacije dokumenta, do generisanja AI/RAG odgovora zasnovanog na tom dokumentu kao validiranom izvoru. Posebno se proverava da sistem ne koristi neaktivne i arhivirane dokumente kao izvore, da prikazuje izvore uz sadržinski odgovor i da ne generiše sadržinski odgovor kada nema dovoljno pouzdanih izvora.

Za funkcionalnost analize korisnički unetog dokumenta faza se prihvata ako sistem uspešno demonstrira upload dokumenta od strane krajnjeg korisnika, proveru formata i limita, analizu dokumenta u odnosu na validiranu bazu znanja, prikaz rezultata analize i izvora, kao i da korisnički uneti dokument nije uključen u validiranu bazu znanja niti dostupan kao izvor drugim korisnicima.

4.4. Faza 4 – Završena aplikacija spremna za UAT

Faza obuhvata dostavljanje kompletne verzije aplikacije spremne za korisničko testiranje. Kompletna verzija obuhvata sve funkcionalnosti predviđene Funkcionalnom specifikacijom za ovu fazu projekta, uključujući: korisnički chat, nastavak razgovora u okviru predviđenih ograničenja, analizu korisnički unetog dokumenta, pretragu baze znanja, prikaz izvora, istoriju razgovora, feedback korisnika, evidentiranje pitanja bez odgovora, osnovnu eskalaciju, admin panel, upravljanje korisnicima, upravljanje dokumentima, metapodacima, statusima i tagovima, token evidenciju, osnovnu analitiku, export podataka i osnovni audit log.

Kriterijum prihvatanja

Faza se prihvata nakon funkcionalnog i performansnog testiranja kojim se potvrđuje da su sve ugovorene funkcionalnosti dostupne u testnom okruženju i spremne za UAT. Posebno se proveravaju: korisnički tok postavljanja pitanja, tok analize korisnički unetog dokumenta, tok pretrage baze znanja, tok unosa i aktivacije dokumenta od strane administratora sadržaja, tok deaktivacije/arhiviranja dokumenta, feedback, pitanja bez odgovora, osnovna eskalacija, osnovna analitika, export podataka, audit log, role i prava pristupa.

Ispravljene kritične greške i prihvaćena testna verzija.

QA, bug fixing, sigurnosne provere, performanse, priprema produkcije

Kriterijum prihvatanja

UAT se smatra uspešno završenim kada su otklonjene sve kritične i visoko prioritetne greške koje utiču na ugovorene funkcionalnosti, bezbednost, korišćenje validiranih izvora, prava pristupa, token evidenciju, rad sa dokumentima, analizu korisnički unetog dokumenta i osnovni rad sistema. Manje greške koje ne utiču na osnovnu funkcionalnost mogu biti evidentirane u zapisniku o uslovnom prihvatanju, uz rok za njihovo otklanjanje

4.6. Faza 6 – Go-live i finalna primopredaja

Cilj faze je instalacija zaokruženog sistema na infrastrukturi SKGO i prenos znanja potreban za samostalno operativno upravljanje. Producersko puštanje sistema, predaja kompletnog izvornog koda, migracija baze podataka, deployment skripti, konfiguracionih šablona, container-ready paketa, tehničke, operativne, administratorske i korisničke dokumentacije, obuka korisnika i administratora i finalna primopredaja sistema

Kriterijum prihvatanja

Završni prijem se vrši nakon produkcionog puštanja sistema, isteka perioda stabilizacije, potvrde SKGO da su otklonjeni nedostaci koji utiču na ugovorene funkcionalnosti i redovan rad sistema, kao i nakon predaje celokupne ugovorene dokumentacije i tehničkih isporuka.

Uslov za prihvatanje Faze 6 je da izvođač preda najmanje: kompletan izvorni kod aplikacije, migracije baze podataka, deployment skripte, konfiguracione šablone, container-ready pakete, spisak komponenti trećih lica sa licencama, tehničku dokumentaciju sistema, dokumentaciju baze podataka, dokumentaciju API-ja, operativno uputstvo, administratorsko uputstvo, korisničko uputstvo i uputstvo za promenu AI providera i embeddings rešenja. Izvođač je dužan da sprovede obuku predstavnika SKGO za operativno korišćenje i osnovno administriranje sistema.

5. Opšti postupak prihvatanja faze

Za svaku fazu primenjuje se sledeći postupak:

Korak	Aktivnost
1. Dostavljanje isporuka	Izvođač dostavlja proizvode faze, prateću dokumentaciju i kratak izveštaj o realizaciji.
2. Provera SKGO	SKGO vrši pregled dokumentacije i testiranje funkcionalnosti u testnom okruženju.
3. Lista nedostataka	SKGO dostavlja izvođaču listu utvrđenih nedostataka i rok za njihovo otklanjanje.
4. Korekcije	Izvođač otklanja nedostatke i dostavlja korigovanu verziju.
5. Prihvatanje	SKGO potpisuje zapisnik o prihvatanju faze ili odobrava uslovno prihvatanje za manje nedostatke koji ne utiču na osnovnu funkcionalnost.
6. Plaćanje	Rata se plaća nakon potpisivanja zapisnika o prihvatanju faze.

Za svaku fazu izvođač je dužan da, uz proizvode faze, dostavi i kratku matricu usklađenosti u kojoj se navodi koje funkcionalnosti iz Funkcionalne specifikacije su isporučene u toj fazi, na koji način su testirane i kojim dokazom se potvrđuje njihova isporuka. SKGO može odbiti prihvatanje faze ako isporuka ne obuhvata funkcionalnosti koje su prema ovom Aneksu i Funkcionalnoj specifikaciji predviđene kao rezultat te faze.

6. Upravljanje promenama obima

Svaka promena u usvojene detaljne tehničke specifikacije mora biti evidentirana kroz pisani zahtev za izmenu od strane SKGO (change request). Zahtev sadrži najmanje: opis izmene, razlog, procenu angažovanja, cenu, uticaj na rokove, uticaj na druge funkcionalnosti i uslove prihvatanja.

Izvođač ne može obračunavati dodatne radove bez prethodnog pisanog odobrenja SKGO.

7. Testiranje i dokazivanje usklađenosti

Izvođač mora da pripremi plan testiranja i sprovede najmanje:

- funkcionalno testiranje svih ugovorenih funkcionalnosti;
- regresiono testiranje nakon korekcija;
- testiranje životnog ciklusa dokumenta;
- testiranje fallback logike;
- testiranje korišćenja neaktivnih i arhiviranih dokumenata;
- testiranje prikaza izvora;
- testiranje evidencije token potrošnje;
- testiranje rola i prava pristupa;
- testiranje performansi i opterećenja;
- osnovni bezbednosni pregled.

Za evaluaciju kvaliteta AI/RAG funkcionalnosti koristi se testni set od najmanje 50 pitanja sa unapred označenim očekivanim izvorima, uključujući pitanja na koja sistem ne sme da generiše sadržinski odgovor. Uzorak odgovora pregleda domenski ekspert kojeg odredi SKGO.

Plan testiranja mora posebno obuhvatiti testove za: životni ciklus dokumenta u SUOD-u, analizu korisnički unetog dokumenta, sprečavanje korišćenja korisnički unetog dokumenta kao izvora za druge korisnike, korišćenje samo aktivnih dokumenata kao izvora, fallback logiku, konflikt izvora, prikaz izvora, pretragu baze znanja, evidenciju token potrošnje, feedback, pitanja bez odgovora, osnovnu eskalaciju, export podataka, audit log, role i prava pristupa

8. Rokovi realizacije

Rok za realizaciju projekta iznosi do **30 nedelja od dana početka projekta**. Danom početka projekta smatra se dan kada je potpisan ugovor.

Izvođač se obavezuje da u sledećim rokovima izvrši faze:

Aplikacija spremna za UAT: u roku od 26 nedelja od započinjanja posla

Produkcioni Go-live: u roku od 30 nedelja od započinjanja posla

Rokovi mogu biti menjani uz saglasnost naručioca i izvođača.

Preduslovi za ispunjenje rokova:

Rokovi realizacije zavise od blagovremene dostupnosti predstavnika Naručioca odnosno pravovremenog dostavljanja potrebne dokumentacije, donošenja odluka, davanja saglasnosti, obezbeđivanja pristupa sistemima, infrastrukturi i drugim resursima koji su izvođaču potrebni za realizaciju projekta.

Nakon isporuke planiranih ishoda pojedinačne faze, Naručilac će u roku od **5 radnih dana** dostaviti pisane primedbe. Ukoliko SKGO u navedenom roku ne dostavi pisane primedbe, smatraće se da je milestone prihvaćen.

Primedbe mogu obuhvatiti samo odstupanja od prethodno usvojene specifikacije. Novi zahtevi, izmene funkcionalnosti ili proširenja obima neće se smatrati nedostacima isporuke, već će biti predmet posebne procedure zahteva za promenom (change request).

U slučaju zahteva SKGO za produženje rokova za dostavljanje pisanih primedbi ili neophodnih informacija, dokumentacije, pristupa, rokovi realizacije se automatski produžavaju za period trajanja takvog kašnjenja, uvećan za vreme potrebno za ponovno planiranje i uključivanje aktivnosti u razvojni ciklus.

Zahtevi za promenom usvojene funkcionalne i tehničke specifikacije

Svi zahtevi koji nisu eksplicitno obuhvaćeni usvojenom funkcionalnom i tehničkom specifikacijom, kao i naknadne izmene poslovnih pravila, korisničkih tokova, integracija, izveštaja, AI pravila, načina indeksiranja dokumenata ili kriterijuma evaluacije odgovora, tretiraće se kao dodatni zahtevi.

Dodatni zahtevi biće predmet posebne procene uticaja na obim, rokove i cenu projekta i sprovodiće se samo nakon saglasnosti obe strane.