

1. Prezentare generală

VOGO eBS AI Agentic Solutions este modulul de inteligență artificială integrat nativ în platforma VOGO eBS, conceput pentru a automatiza comunicarea, asista utilizatorii în timp real și orchestra fluxuri de lucru complexe în medii enterprise. Soluția combină LLM-uri de ultimă generație cu arhitecturi agentic AI — RAG, LangChain, LangGraph și MCP — oferind organizațiilor capabilități AI avansate fără dependență de platforme terțe.

2. Arhitectura tehnică & Stivă LLM

Provideri LLM suportați

Platforma suportă configurarea și comutarea în timp real între mulți provideri LLM, fără modificări de cod:

- ▶ Anthropic (Claude) — raționament avansat, context extins, securitate enterprise
- ▶ OpenAI (GPT-4o, GPT-4 Turbo) — model de referință pentru aplicații conversaționale
- ▶ Google Gemini — integrare nativă cu ecosistemul Google Workspace
- ▶ Ollama — rulare modele open-source (LLaMA, Mistral, Qwen) local, on-premise, fără costuri per token
- ▶ Groq — inferență ultra-rapidă pentru scenarii cu latență critică
- ▶ xAI (Grok) — model alternativ pentru use-case-uri specifice
- ▶ Mistral — model european, relevant pentru conformitate GDPR strict
- ▶ OpenRouter — gateway unificat pentru acces la zeci de modele prin un singur API

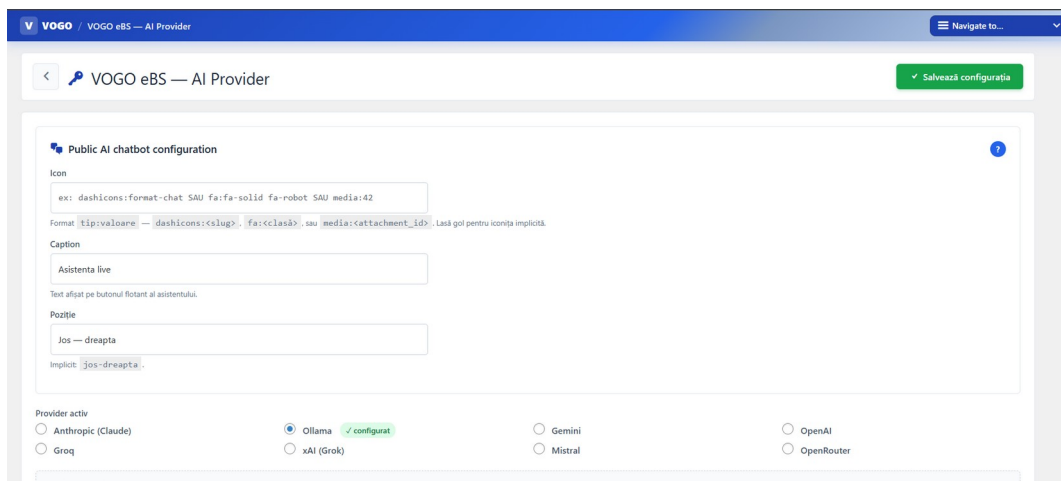


Fig. 1 — Configurarea providerului AI activ din interfața VOGO eBS AI Provider

Componente arhitecturale

- ▶ LangChain — abstractizare provider LLM, RAG/Retrievers, Output Parsers, Tool Use
- ▶ LangGraph — orchestrare fluxuri agentic cu stare (state machines), bucle, condiții, retry logic, Human-in-the-Loop
- ▶ LangSmith — observabilitate completă: tracing apeluri, testing/evals, monitoring producție, versionare prompturi
- ▶ RAG (Retrieval-Augmented Generation) — baze de cunoștințe proprii fără re-training; retrieval top-K chunks → răspuns contextual precis
- ▶ MCP (Model Context Protocol) — standard deschis pentru conectarea agenților la tool-uri și surse externe (ERP, CRM, DMS, calendare); Embeddings + Vector Database — indexare semantică a documentelor (Pinecone, ChromaDB, pgvector) pentru căutare contextuală instantă

3. Capacități funcționale

AI Live Assistant — Asistent integrat în formulare

Platforma include un asistent AI integrat nativ în modulul de Forms care ghidează utilizatorii în crearea și completarea formularelor digitale în timp real.

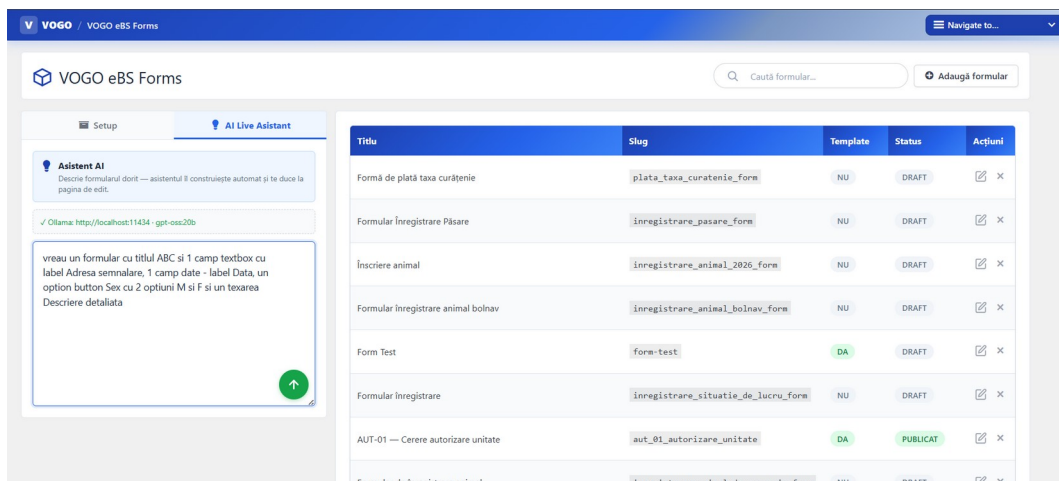


Fig. 2 — AI Live Assistant activ în modulul VOGO eBS Forms

AI Live Support — Suport inteligent în servicii

Modulul de servicii beneficiază de un strat AI de suport live care asistă operatorii și cetățenii în navigarea fluxurilor de servicii digitale, reducând timpii de rezolvare și volumul de solicitări manuale.

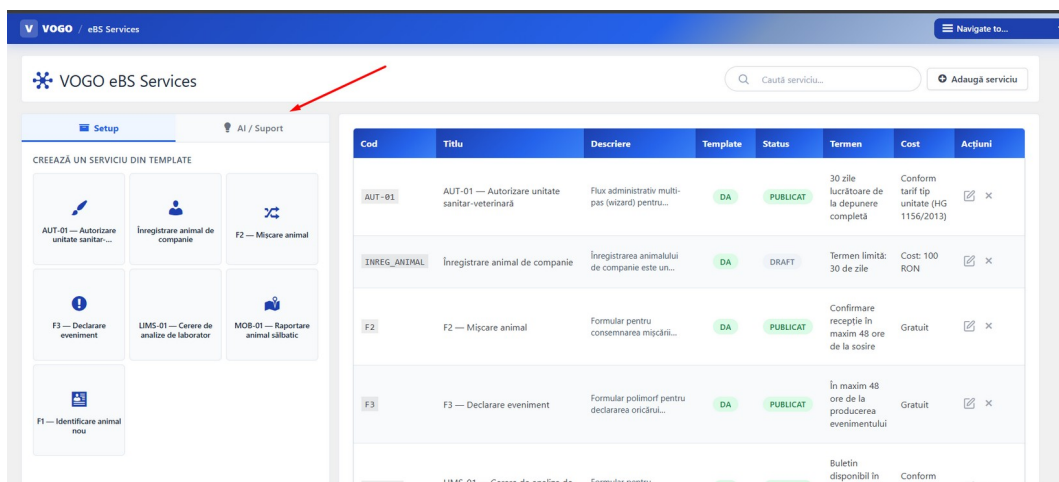


Fig. 3 — AI Live Support activ în modulul VOGO eBS Services

Chatbot & Automatizare comunicare

- ▶ Chatbot bazat pe AI conversațional cu recunoaștere text și voce, comportament asemănător uman
- ▶ Funcționare integrală on-premise, fără dependență de servicii cloud externe
- ▶ Modul NLU (Natural Language Understanding) proprietar pentru recunoașterea intențiilor și entităților
- ▶ Modul Q&A configurabil de administratori fără scriere de cod
- ▶ Mesaje de întâmpinare personalizat trimis automat la deschiderea sesiunii de chat
- ▶ Aplicație web pentru canalul chat, integrabilă în orice pagină HTML
- ▶ Fluxuri Human-in-the-Loop (HITL): agentul AI propune, operatorul uman aprobă sau intervine

Canale de comunicare suportate

- ▶ Chat web — widget embeddabil în orice pagină HTML
- ▶ E-mail — procesare și răspuns automat
- ▶ SMS — notificări și interacțiuni bidireționale
- ▶ Social media — integrare configurabilă de administrator (Facebook, WhatsApp și altele)
- ▶ Arhitectură multi-canal extensibilă pentru canale adiționale

Management utilizatori & SSO

- ▶ Moduri diferențiate de interacțiune pentru utilizatori interni și externi

- ▶ Cont extern anonim (bazat pe cookie browser) sau autentificat cu user/parolă; Sincronizare cu sistemul SSO existent — utilizatorii interni nu gestionează conturi multiple; Creare proiect chatbot nou disponibilă administratorului principal al platformei
- ▶ Administratorul de proiect configurează integrările cu sisteme externe și canale sociale

4. Capabilități AI avansate

Procesare & extragere documente

- ▶ Ingestie automată PDF, Word, HTML, CSV, Notion, Confluence prin LangChain Document Loaders
- ▶ OCR pentru documente scanate (Tesseract, AWS Textract, Google Document AI)
- ▶ Chunking inteligent: RecursiveCharacterTextSplitter, TokenTextSplitter, SemanticChunker
- ▶ Extragere metadata pentru filtrare și retrieval contextual în sisteme enterprise cu mii de documente

Memorie & Cache

- ▶ Conversation Memory (Buffer, Window, Summary) — persistența contextului în chatboți
- ▶ Semantic Memory — amintiri relevante din sesiuni anterioare pentru personalizare
- ▶ Prompt Caching — discount 90% (Anthropic) / 75% (Google) pe system prompt-uri repetate
- ▶ LangGraph Checkpointing — memorie persistentă între sesiuni fără cod custom

Optimizare costuri & performanță

- ▶ Router patterns — query-uri simple → modele ieftine; query-uri complexe → modele premium (60-80% reducere costuri)
- ▶ Async & Batching — procesare paralelă, Anthropic Batch API (50% discount față de real-time)
- ▶ Quantization (GGUF, GPTQ, AWQ) — rulare modele mari pe hardware standard
- ▶ Structured Output / Pydantic — răspunsuri JSON validate automat, fără parsing fragil

5. Securitate & Conformitate

- ▶ PII Detection & Redaction (Presidio, AWS Comprehend) — mascarea automată date personale înainte de trimitere la LLM cloud
- ▶ Guardrails (NeMo Guardrails, LlamaGuard) — reguli de comportament configurabile: topice interzise, detectare conținut toxic
- ▶ Protecție Prompt Injection — sanitizare input, validare output, system prompt robust
- ▶ Rate Limiting & Auth — autentificare per user, limitare RPM/TPM, audit log complet
- ▶ Output validation pipelines — scoring încredere, detectare halucinații, verificare surse
- ▶ Conformitate GDPR — date procesate local (Ollama on-premise), fără exfiltrare date

6. Opțiuni de Deployment

Cloud (SaaS)	Deploy managed pe infrastructură cloud; scalare automată, zero management server
On-Premise	Instalare pe infrastructura proprie; date nu părăsesc organizația; ideal GDPR strict
Hybrid	LLM local (Ollama) + orchestrare cloud; echilibru cost-conformitate
SageMaker (AWS)	Deployment modele fine-tuned pe AWS; autoscaling, Spot Instances, cost optimizat
Docker / Lambda	Arhitectură serverless; Lambda + SageMaker endpoint; fără servere de gestionat

7. Integrări suportate

- ▶ VOGO Portal — asistent AI disponibil direct în portalul organizației
- ▶ VOGO Forms — AI Live Assistant pentru crearea și completarea formularelor
- ▶ VOGO Services — AI Live Support pentru fluxuri de servicii digitale
- ▶ DMS (Document Management System) — acces la documente prin servicii web
- ▶ CRM, ERP, Ticketing — integrare prin MCP sau API REST
- ▶ SAP, Oracle eBS, Microsoft D365 — conectori enterprise pentru date operaționale
- ▶ E-mail, SMS, Social Media — canale de comunicare multi-canal

8. Alte functionalitati

- ▶ Modulul contine un sistem de automatizare a comunicării cu cetățenii — asistent AI disponibil care poate fi afigisat atat direct în portalul organizației cat si aplicatia mobila corespunzatoare (Androis sau iOS) care poate fi publicata in Google play si Apple Store
- ▶ Chatbot, bazat pe inteligență artificială, programat să desfășoare conversații cât mai asemănătoare comportamentului uman, prin recunoaștere text și voce.
- ▶ Poate sa funcționeze integral on-premise, fără dependență de servicii externe.
- ▶ Oferă capabilitatea de integrare cu multiple canale de comunicare.
- ▶ Pentru canalul de chat web, pune la dispoziție o aplicație web integrabilă în orice pagină HTML.
- ▶ Permite administratorului de proiect să activeze și să configureze integrări cu sisteme externe și canale de social media.
- ▶ Permite configurarea și administrarea conținutului asistentului în mod facil, fără scriere de cod.
- ▶ Pune la dispoziție instrumente complete pentru testarea chatboților: testarea fluxului conversațional, a intențiilor și a entităților recunoscute.
- ▶ Permite interacțiunea cu soluția DMS prin intermediul serviciilor web ale acesteia.
- ▶ Permite integrarea cu soluții de e-mail și SMS, cu posibilitate de extindere pentru canale adiționale.
- ▶ Permite moduri diferite de interacțiune pentru utilizatorii interni și externi.
- ▶ Permite fluxuri de tipul „Human in the Loop” (HITL).
- ▶ Permite sincronizarea cu sistemul SSO, eliminând necesitatea gestionării multiple a conturilor pentru utilizatorii interni.
- ▶ Include un modul LLM dedicat.
- ▶ Conține un modul proprietar de înțelegere a limbajului natural (NLU), bazat pe inteligență artificială, care recunoaște intențiile și entitățile exprimate liber de utilizatori.
- ▶ Permite configurarea unui mesaj de întâmpinare trimis automat la deschiderea ferestrei de chat. Permite configurarea de intrebari si raspunsuri predevinite, ierarhice, pe model de script conversational
- ▶ Permite integrarea cu soluțiile LLM consacrate cum ar fi dar fara a ne limita la: LLaMA, OpenAI, Gemini.
- ▶ Permite fiecărui utilizator extern să creeze un cont anonim (bazat pe cookie de browser) sau autentificat cu user și parolă proprie.