# Архитектурная спецификация: Универсальный AI-агрегатор услуг

## 1. Общее описание

Данное архитектурное предложение представляет собой гибридную AI-платформу, сочетающую облачные и edge-компоненты.   
Платформа предназначена для агрегации различных услуг (интернет, финансы, коммунальные, юридические и др.) с использованием AI-инструментов: чат-ботов, семантического поиска, аналитики, KYC/AML проверок и антифрода.

## 2. Принципы архитектуры

* • Минимизация затрат: использование open-source и edge-инференса
* • Гибкость и масштабируемость: microservices, cloud + edge
* • Безопасность и конфиденциальность: client-side шифрование, дифференциальная приватность
* • AI-first подход: автономность, генерация и обработка контента
* • Модульность и Lego-style расширяемость
* • SEO-оптимизация и SSR
* • Event-driven интеграции и Webhook-инфраструктура

## 3. Компоненты системы

* Frontend:

Next.js, Tailwind, PWA, i18n, SSR/SEO, интеграция с Edge Functions

* Backend API:

FastAPI, LangChain, Redis (Upstash), PostgreSQL (Supabase JSONB + pgvector)

* AI-инференс:

Llama.cpp, Mixtral 8x22B (локально на Edge), Hugging Face API, Google LLM Inference API

* Semantic Search:

ChromaDB + pgvector + LangChain

* DevOps:

Docker, GitHub Actions, Tekton/ArgoCD, Prometheus + Grafana

* Security:

Cloudflare WAF, TensorFlow Privacy, Web Crypto API, шифрование на клиенте

* Integration:

Webhooks, Telegram/Slack боты, Event-driven архитектура

* CRM и аналитика:

Supabase dashboard + кастомные метрики + Edge Analytics

* Мультиязычность:

i18n, Supabase RLS, white-label брендинг

## 4. Безопасность и соответствие

Система проектируется с учётом GDPR, PCI DSS и best practices OWASP:  
• KYC/AML проверки поставщиков и пользователей  
• Client-side encryption и Edge-inference минимизируют утечку данных  
• Web Crypto API и Firebase AppCheck (опционально)  
• Дифференциальная приватность при сборе аналитики

## 5. Заключение

Предложенная архитектура обеспечивает гибкость, отказоустойчивость, масштабируемость и экономическую эффективность.   
Edge-инференс и open-source подходы значительно повышают независимость и безопасность, а AI-first принципы — скорость и интеллектуальность системы.