**ecnologias do Chat Energy - Análise Completa**

| **Categoria** | **Tecnologia** | **Versão** | **Finalidade** | **✅ Vantagens** | **❌ Desvantagens** |
| --- | --- | --- | --- | --- | --- |
| **🐍 Backend Core** | Python | 3.12 | Linguagem principal | • Sintaxe clara e legível<br>• Vasta comunidade<br>• Excelente para IA/ML<br>• Libraries abundantes | • Performance inferior a linguagens compiladas<br>• GIL limita paralelismo<br>• Tipagem dinâmica pode gerar bugs |
|  | FastAPI | ≥0.104.0 | Framework web assíncrono | • Alta performance<br>• Documentação automática<br>• Type hints nativo<br>• Assíncrono por padrão | • Curva de aprendizado<br>• Menos maduro que Django<br>• Comunidade menor |
|  | Uvicorn | ≥0.24.0 | Servidor ASGI | • Muito rápido<br>• Suporte WebSocket<br>• Baixo uso de memória | • Configuração complexa<br>• Debugging difícil<br>• Menos recursos que Gunicorn |
| **🤖 Inteligência Artificial** | OpenAI | ≥1.0.0 | Modelo LLM principal | • Estado da arte em qualidade<br>• API simples<br>• Documentação excelente<br>• Múltiplos modelos | • Custo elevado<br>• Dependência externa<br>• Latência de rede<br>• Rate limits |
|  | LangChain | ≥0.1.0 | Framework RAG | • Abstração poderosa<br>• Integração fácil<br>• Chains complexas<br>• Comunidade ativa | • Overhead de performance<br>• Mudanças frequentes<br>• Debugging complexo<br>• Documentação inconsistente |
|  | Tiktoken | ≥0.5.0 | Tokenização OpenAI | • Precisão perfeita<br>• Otimizado para OpenAI<br>• Rápido | • Específico para OpenAI<br>• Limitado a alguns modelos<br>• Dependência extra |
| **🔍 RAG & Embeddings** | FAISS | ≥1.7.4 | Banco vetorial | • Performance excepcional<br>• Escalabilidade<br>• GPU support<br>• Facebook backing | • Complexidade de configuração<br>• Uso intensivo de memória<br>• Curva de aprendizado |
|  | Sentence Transformers | ≥2.2.2 | Modelos de embedding | • Modelos pré-treinados<br>• Multilingual support<br>• Fácil de usar | • Modelos grandes<br>• RAM intensivo<br>• GPU recomendada |
|  | HuggingFace | - | Modelo multilingual | • Suporte 100+ idiomas<br>• Qualidade alta<br>• Gratuito | • Tamanho do modelo (1.5GB)<br>• Inicialização lenta<br>• RAM intensivo |
| **🗄️ Banco de Dados** | SQLite | - | Banco principal | • Zero configuração<br>• Portável<br>• Rápido para reads<br>• Embarcado | • Sem concorrência de writes<br>• Limitações de escala<br>• Não distribuído |
|  | SQLAlchemy | ≥2.0.0 | ORM | • ORM poderoso<br>• Suporte multi-DB<br>• Migrations automáticas | • Overhead de performance<br>• Curva de aprendizado<br>• Queries complexas |
| **🔐 Autenticação** | Passlib | ≥1.7.4 | Hash de senhas | • Múltiplos algoritmos<br>• Seguro por padrão<br>• Salt automático | • Configuração manual<br>• Dependências extras |
|  | JWT | ≥2.8.0 | Tokens | • Stateless<br>• Escalável<br>• Padrão da indústria | • Não pode ser revogado<br>• Payload visível<br>• Vulnerável se mal implementado |
| **📄 Processamento** | PyPDF | ≥3.0.0 | Leitura de PDFs | • Gratuito<br>• Simples<br>• Suporte OCR | • PDFs complexos falham<br>• Limitações de formatação<br>• Não processa imagens |
|  | Python-docx | ≥0.8.11 | Documentos Word | • API intuitiva<br>• Suporte completo DOCX | • Apenas DOCX moderno<br>• Não suporta DOC legado<br>• Formatação limitada |
| **🌐 Frontend** | HTML5 | - | Estrutura | • Padrão universal<br>• Semântica rica<br>• SEO friendly | • Verboso<br>• Não dinâmico<br>• Requer JavaScript |
|  | Tailwind CSS | CDN | Framework CSS | • Utility-first<br>• Customização total<br>• Performance | • Curva de aprendizado<br>• HTML verboso<br>• Bundle size |
|  | JavaScript ES6+ | - | Interatividade | • Linguagem universal<br>• Ecossistema rico<br>• Evolução constante | • Tipagem fraca<br>• Callback hell<br>• Compatibilidade |
| **🔗 Comunicação** | WebSockets | ≥12.0 | Chat tempo real | • Baixa latência<br>• Bi-direcional<br>• Eficiente | • Complexidade de estado<br>• Proxy issues<br>• Debugging difícil |
|  | HTTPX | ≥0.25.0 | Cliente HTTP | • Async/await nativo<br>• HTTP/2 support<br>• API moderna | • Menos maduro<br>• Documentação limitada<br>• Breaking changes |
| **⚙️ Configuração** | Python-dotenv | ≥1.0.0 | Variáveis ambiente | • Simples de usar<br>• Padrão da indústria<br>• Segurança | • Arquivo não versionado<br>• Sincronização manual<br>• Sem validação |
| **🧪 Testes** | Pytest | ≥7.4.0 | Framework testes | • Sintaxe simples<br>• Plugins abundantes<br>• Fixtures poderosas | • Configuração complexa<br>• Debugging difícil<br>• Performance |
| **☁️ Deploy** | Docker | - | Containerização | • Portabilidade<br>• Isolamento<br>• Escalabilidade | • Overhead de recursos<br>• Curva de aprendizado<br>• Complexidade de rede |
|  | Google Cloud Run | - | Hospedagem | • Serverless<br>• Auto-scaling<br>• Pay-per-use<br>• Zero config | • Cold starts<br>• Timeouts (15min)<br>• Vendor lock-in<br>• Limitações de estado |
|  | Nginx | - | Servidor web | • Performance alta<br>• Configuração flexível<br>• Load balancing | • Configuração complexa<br>• Sintaxe específica<br>• Debugging difícil |