

Requisitos Detalhados para o MVP

Visão Geral

Este documento detalha os requisitos funcionais e não-funcionais para o Produto Mínimo Viável (MVP) da plataforma de atendimento ao cliente automatizado com IA.

Funcionalidades Essenciais

1. Integração com WhatsApp

- Conectar com WhatsApp Business API
- Receber e processar mensagens de texto
- Enviar respostas de texto
- Capturar metadados das conversas

2. Processamento de Documentos

- Upload de documentos (PDF, DOCX, TXT)
- Processamento via docling (Python)
- Extração de texto e metadados
- Chunk de texto para indexação
- Geração de embeddings

3. Motor RAG (Retrieval Augmented Generation)

- Indexação de conhecimento em vector database
- Busca semântica por informações relevantes
- Contextualização do LLM com informações recuperadas
- Geração de resposta contextualizada

4. Dashboard Administrativo

- Login e gestão de usuários
- Visualização de métricas básicas
- Gerenciamento de documentos
- Histórico de conversas
- Gestão de créditos

5. Sistema de Créditos

- Controle de saldo por cliente

- Débito automático por uso
- Compra e adição de créditos
- Alertas de saldo baixo

6. Sistema de Fallback

- Detecção de baixa confiança
- Notificação de casos não resolvidos
- Mensagens padronizadas para temas fora do escopo

User Stories Prioritárias

Para Clientes (Empresas)

Onboarding e Configuração

- Como cliente, quero criar uma conta para acessar o sistema
- Como cliente, quero conectar meu número de WhatsApp Business para atendimento
- Como cliente, quero personalizar o nome e tom de voz do assistente
- Como cliente, quero definir mensagens de boas-vindas e despedida

Gestão de Conhecimento

- Como cliente, quero fazer upload de documentos para treinar meu assistente
- Como cliente, quero ver o status do processamento dos meus documentos
- Como cliente, quero adicionar FAQs manualmente para casos específicos
- Como cliente, quero organizar meus documentos com tags e categorias

Monitoramento e Análise

- Como cliente, quero visualizar histórico de conversas
- Como cliente, quero ver métricas básicas de atendimento (volume, resolução)
- Como cliente, quero identificar tópicos frequentes nas conversas
- Como cliente, quero receber alertas sobre problemas recorrentes

Gestão de Créditos

- Como cliente, quero visualizar meu saldo atual de créditos
- Como cliente, quero comprar créditos adicionais
- Como cliente, quero ver meu histórico de consumo
- Como cliente, quero receber alertas quando meu saldo estiver baixo

Para Usuários Finais

- Como usuário final, quero iniciar uma conversa pelo WhatsApp facilmente
- Como usuário final, quero obter respostas rápidas e precisas
- Como usuário final, quero ser transferido para um humano quando necessário
- Como usuário final, quero receber confirmação de que estou falando com um assistente virtual

Para Administradores do Sistema

- Como administrador, quero monitorar o uso geral da plataforma
- Como administrador, quero gerenciar contas de clientes
- Como administrador, quero visualizar métricas globais
- Como administrador, quero configurar preços e pacotes de créditos

Requisitos Não-Funcionais

Desempenho

- Tempo de resposta do assistente < 5 segundos
- Capacidade de processar até 100 conversas simultâneas por cliente
- Processamento de documentos concluído em até 10 minutos por documento (50 páginas)
- Tempo de carregamento do dashboard < 3 segundos

Segurança

- Autenticação multi-fator para dashboard
- Criptografia de dados em trânsito e repouso
- Isolamento de dados entre clientes
- Logs de auditoria para ações sensíveis

Disponibilidade

- Uptime de 99.5% para MVP
- Janela de manutenção programada fora do horário comercial
- Backup diário de dados

Escalabilidade

- Arquitetura que suporte crescimento horizontal
- Componentes stateless quando possível
- Preparado para aumentar capacidade com demanda

Usabilidade

- Interface intuitiva para dashboard
- Design responsivo (desktop e mobile)
- Documentação de ajuda contextual
- Onboarding guiado para novos clientes

Fluxos Principais

1. Fluxo de Configuração Inicial

1. Cliente cria conta no sistema
2. Cliente configura perfil do assistente (nome, tom, boas-vindas)
3. Cliente conecta WhatsApp Business
4. Cliente faz upload de documentos iniciais
5. Sistema processa documentos (docling + embeddings)
6. Cliente compra créditos iniciais
7. Cliente ativa o assistente

2. Fluxo de Atendimento

1. Usuário envia mensagem via WhatsApp
2. Sistema recebe e processa a mensagem
3. Sistema consulta base de conhecimento (RAG)
4. Sistema gera resposta contextualizada
5. Resposta é enviada ao usuário
6. Sistema debita créditos da conta
7. Conversa é registrada para análise

3. Fluxo de Análise

1. Cliente acessa dashboard
2. Cliente visualiza métricas de atendimento
3. Cliente identifica tópicos frequentes
4. Cliente explora conversas específicas
5. Cliente atualiza base de conhecimento conforme necessidade

Limitações do MVP

O que está incluído

- Atendimento via WhatsApp apenas
- Respostas de texto (sem imagens ou áudio)
- Português como idioma primário
- Processamento básico de documentos (texto)
- Métricas essenciais de atendimento
- Sistema de pagamento manual/básico

O que não está incluído

- Integração com outros canais (será adicionado após MVP)
- Respostas com imagens ou áudio (será adicionado após MVP)
- Múltiplos idiomas (será adicionado após MVP)
- Processamento avançado de tabelas e imagens (será adicionado após MVP)
- IA especializada para SDR e vendas (será adicionado após MVP)
- Integração com sistemas CRM (será adicionado após MVP)

CrITÉrios de Aceitação

Integração com WhatsApp

- Mensagens são recebidas corretamente em até 5 segundos
- Respostas são enviadas corretamente em até 5 segundos
- Sistema identifica corretamente o usuário
- Metadados da conversa são salvos adequadamente

Processamento de Documentos

- Documentos DOCX, PDF e TXT são processados corretamente
- Texto é extraído com precisão > 95%
- Embeddings são gerados e indexados corretamente
- Documentos podem ser organizados e gerenciados

Mecanismo de RAG

- Sistema retorna contextos relevantes para a query
- Respostas são geradas usando o contexto recuperado
- Taxa de resposta correta > 70% em teste controlado
- Casos de baixa confiança são identificados

Dashboard

- Todas as métricas são exibidas corretamente
- Histórico de conversas é acessível e navegável
- Interface é responsiva em diferentes dispositivos
- Controles de gerenciamento funcionam corretamente

Sistema de Créditos

- Saldo é atualizado corretamente após cada interação
- Compra de créditos é processada corretamente
- Alertas de saldo baixo são enviados
- Histórico de consumo é preciso e detalhado

Métricas de Sucesso do MVP

Métricas Técnicas

- Tempo de resposta < 5 segundos em 95% das interações
- Uptime > 99.5%
- Taxa de erros técnicos < 5%
- Completude de processamento de documentos > 95%

Métricas de Negócio

- Taxa de resolução automática > 60%
- Satisfação do usuário > 4/5 (medida via feedback)
- Conversão de clientes piloto para pagantes > 50%
- Retenção após 30 dias > 80%

Métricas de Produto

- Uso ativo > 50 conversas/mês por cliente
- Crescimento no volume de créditos adquiridos mês a mês > 10%
- Tempo médio para setup completo < 2 dias
- Taxa de adoção de features > 70%

Dependências e Integrações

Dependências Externas

- WhatsApp Business API
- API de LLM (GPT/Claude)
- Serviços de pagamento

- Infraestrutura de nuvem

Integrações Internas

- Serviço Node.js <-> Serviço Python (docling)
- Vector Database <-> Engine de IA
- Sistema de orquestração <-> Serviços específicos

Riscos e Mitigações

Riscos Técnicos

- **Risco:** Latência alta no processamento de documentos grandes
 - **Mitigação:** Implementar processamento assíncrono e notificações
- **Risco:** Qualidade inconsistente das respostas do LLM
 - **Mitigação:** Implementar mecanismos de filtragem e validação
- **Risco:** Problemas de integração com WhatsApp Business API
 - **Mitigação:** Testes extensivos e plano de contingência

Riscos de Negócio

- **Risco:** Custo elevado das APIs de LLM
 - **Mitigação:** Monitoramento de uso e otimização de prompts
- **Risco:** Baixa adoção por clientes
 - **Mitigação:** Programa de onboarding assistido e feedback contínuo
- **Risco:** Concorrência de soluções similares
 - **Mitigação:** Foco em diferenciais como suporte ao português e facilidade de uso

Próximos Passos após MVP

Melhorias Imediatas

- Otimização de performance baseada em métricas reais
- Refinamento da UX baseado em feedback de clientes
- Melhorias na qualidade de resposta baseadas em análise de conversas

Features para Próxima Fase

- Integração com canais adicionais (Telegram, web chat)
- Suporte a respostas com imagens e arquivos
- Integração com sistemas CRM populares
- Dashboard avançado de análise de conversas

Apêndices

Modelo de Dados Simplificado

Cliente

- id
- nome
- email
- senha
- saldo_creditos
- configuracoes_assistente
- data_criacao

Documento

- id
- cliente_id
- nome
- tipo
- status
- metadados
- data_upload

Conversa

- id
- cliente_id
- usuario_id
- canal
- status
- data_inicio
- data_fim

Mensagem

- id
- conversa_id
- conteudo
- tipo
- remetente
- timestamp
- metadados

Credito

- id
- cliente_id
- quantidade
- tipo_operacao
- timestamp
- saldo_apos

Stack Tecnológico Recomendado

Frontend:

- React com Next.js
- Tailwind CSS
- Auth0 para autenticação

Backend:

- Node.js com Express
- Python/Flask para serviço de processamento de documentos
- PostgreSQL para dados relacionais
- Pinecone/pgvector para vector database
- Redis para cache
- RabbitMQ para mensageria

Infraestrutura:

- AWS (EC2, S3, RDS)
- Docker para containerização
- GitHub Actions para CI/CD
- Sentry para monitoramento de erros