

Dinâmica Prática III – Planejando o Agente

Agente de IA para Triagem de Chamados TI

Transformando o suporte técnico com inteligência artificial e automação inteligente



Agente Mínimo Viável (MVP)

Agente Baseado em **Modelo**

Identifica o problema priorizando urgência e consome base de conhecimento para identificar resoluções similares baseadas em dados estatísticos históricos.

- Classificação baseada em padrões conhecidos
- Roteamento por regras de prioridade
- Sem capacidade de aprendizado adaptativo
- Resolve aproximadamente 70% dos casos



Limitações do MVP



Por que Evoluir para Fase 2?

O agente MVP apresenta restrições importantes que limitam sua efetividade:

- Incapaz de aprender com feedback operacional em tempo real;
- Dificuldade em identificar padrões complexos e novos tipos de incidentes;
- Não prioriza adequadamente entre múltiplos chamados críticos simultâneos;
- Ausência de capacidade de auto-resolução para problemas simples;
- Histórico e feedback.

Agente Otimizado: Utilidade + Aprendizado com LLM

Combina capacidade de aprendizado contínuo com otimização de função de utilidade para maximizar valor de negócio.

Incorporação de LLM: O Large Language Model atua na construção da base de conhecimentos, detecção de padrões complexos e geração de planos de resolução contextualizados.



Aprendizado e Histórico

Evolui continuamente com base em resoluções passadas e avaliações dos atendentes



Auto-resolução Inteligente de Incidentes

Resolve automaticamente problemas recorrentes sem intervenção humana



Detecção Avançada de Padrões

Identifica tendências e incidentes massivos antes que se tornem críticos



Priorização Contextual Sofisticada

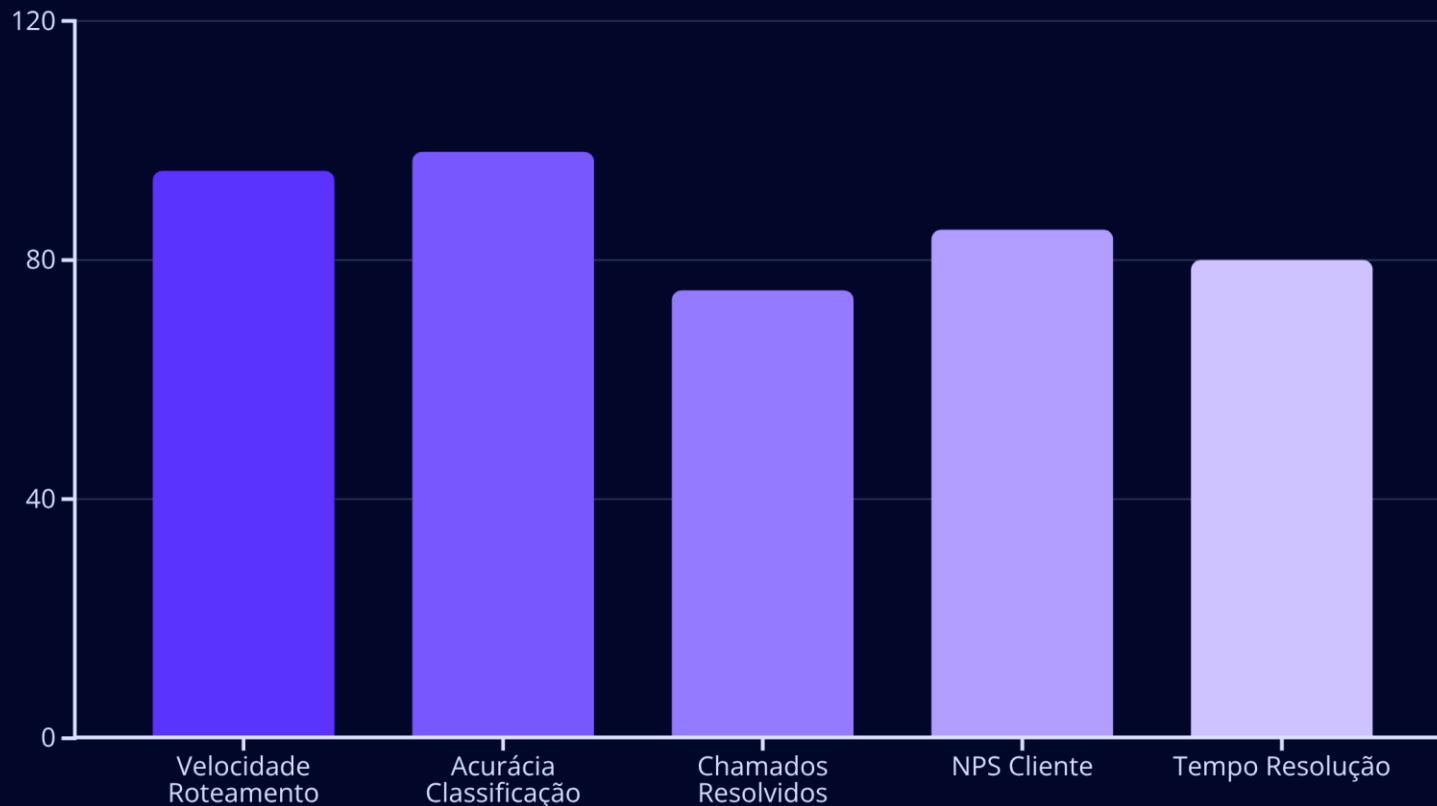
Define a ordem de atendimento considerando urgência, impacto e contexto do negócio

Beliefs: Dados e Regras de Negócio/Compliance

Para operar efetivamente, o agente precisa acessar múltiplas fontes de informação:

Categoria	Dados Específicos
Chamado	Texto, sistema afetado, horário, canal de abertura
Usuário	Nome, cargo, departamento, VIP status, histórico de chamados
Sistemas	Criticidade, time responsável, horário de suporte
Times	Especialidades, disponibilidade, carga atual, SLA
Histórico	Chamados similares, soluções aplicadas, tempo de resolução, incidentes
Base de Conhecimento	Artigos de solução, FAQs, procedimentos, documentação técnica
Incidentes Ativos	Problemas conhecidos em andamento
Políticas de Atendimento	Diretrizes de comunicação, escalação e procedimentos operacionais padrão
Requisitos de Auditabilidade	Trilhas de decisão, logs de operação e conformidade com regulamentações

Função de Utilidade e Métricas



A função de utilidade deve equilibrar todas essas dimensões para otimizar o valor global entregue.

↗ Métricas Primárias

- **Maximizar:** Velocidade de roteamento
- **Minimizar:** Erros de classificação

↗ Métricas Secundárias

- Quantidade de chamados resolvidos
- Satisfação do cliente (NPS)
- Tempo médio de resolução

Raciocínio e Tolerância a Falhas

Sequências de Raciocínio: O LLM precisa gerar (1) classificação com contexto enriquecido, (2) decisão de auto-resolução baseada em confiança, e (3) detecção proativa de incidentes massivos.

Falhas NÃO Aceitáveis

- Exposição de dados sensíveis (violação LGPD)
- Tempo de roteamento superior a 2 minutos
- Classificar chamado "Crítico" como "Baixo"
- Não detectar incidente massivo em andamento

Falhas Aceitáveis

Situações onde imprecisões são toleráveis:

- Confusão entre categorias próximas (ex: "Rede" → "Infraestrutura")
- Solicitar informações adicionais quando há incerteza na classificação

Fatores Críticos de Sucesso



Latência

Resposta em segundos para garantir experiência ágil



Privacidade

Conformidade rigorosa com LGPD e proteção de dados



Disponibilidade

Operação contínua 24/7 sem interrupções



Auditabilidade

Rastreabilidade total de todas as decisões do agente



Assertividade

Alta acurácia na classificação e roteamento



Integração

APIs robustas com sistemas existentes de TI