



ANEXO TÉCNICO: JUSTIFICATIVA DE EFICIÊNCIA E OTIMIZAÇÃO NO CONSUMO DE INTELIGÊNCIA ARTIFICIAL

Referência: Pregão Eletrônico PEE 2026000033

Objeto: Fornecimento e Implantação de Plataforma de Automação e Orquestração de Processos (SaaS)

1. Objetivo

Este documento visa detalhar a arquitetura de Inteligência Artificial nativa da solução proposta, comprovando a capacidade técnica de atendimento aos requisitos do Termo de Referência (leitura de documentos, preenchimento de campos e roteamento inteligente), associada a mecanismos avançados de otimização de consumo de processamento cognitivo (tokens/LLM).

2. Arquitetura do Motor de IA (Portal de aprovações + N8N workflow)

A solução utiliza um modelo híbrido focado em eficiência operacional e escalabilidade estrutural:

- **Orquestração Low-Code/No-Code:** O ecossistema baseado em nós gerencia o fluxo de dados de forma preditiva, garantindo que os modelos de linguagem em grande escala (LLMs) sejam acionados estritamente quando houver real necessidade de tomada de decisão ou análise cognitiva.
- **Camada de Abstração de IA (Portal de Aprovações):** Atua diretamente na governança dos agentes inteligentes através de uma interface intuitiva. Padroniza os prompts estruturados e refina o contexto que é enviado para as APIs de IA, o que reduz drasticamente o desperdício de processamento e o tráfego de dados redundantes.
- **Alta Disponibilidade e Resiliência em Nuvem:** O ecossistema completo é configurado, aplicado e mantido em ambiente de nuvem de alta disponibilidade (AWS / Google Cloud). A arquitetura foi desenhada com camadas de redundância ativa e balanceamento de carga, garantindo operação contínua 100% do tempo e proteção contra desastres.

3. Mecanismos de Otimização e Proteção de Custos

Para suportar o modelo de uso institucional transversal e o volume variável de demandas exigidos pelo Senac SP, a plataforma conta com três travas técnicas de eficiência:

A. Camada de Caching Inteligente de Contexto

A plataforma impede o reprocessamento de dados idênticos. Se um documento ou instrução já foi analisado pela IA dentro de um fluxo de automação, o sistema

recupera a resposta diretamente de um banco de dados de alta velocidade (Cache em memória), eliminando novas chamadas pagas às APIs de LLM.

B. Arquitetura de Roteamento Heurístico antes da IA

Antes de enviar qualquer tarefa para a Inteligência Artificial, o motor de automação aplica regras de negócio e validações heurísticas locais (sem custo de processamento de IA).

- *Exemplo:* Em vez de usar a IA para ler um documento inteiro apenas para saber se ele possui assinatura, filtros lógicos tradicionais validam os metadados do arquivo primeiro. A IA só é acionada para a interpretação contextual complexa do conteúdo.

C. Gestão Contratual de Quotas e Rate Limiting

Considerando a premissa de usuários Solicitantes e Aprovadores ilimitados estabelecida no edital, a plataforma implementa políticas de *Rate Limiting* (limitação de taxa de requisições simultâneas por usuário/unidade). Isso mitiga riscos de loops de automação informáticos ou uso abusivo, blindando a alta disponibilidade do ambiente SaaS.

4. Segurança, Privacidade e Conformidade (LGPD)

Todo o processamento de Inteligência Artificial atende rigorosamente aos critérios de segurança exigidos:

- **Isolamento de Dados:** Os dados institucionais transacionados nos fluxos do Senac SP não são utilizados para o treinamento público de modelos de IA de terceiros, garantindo sigilo absoluto.
- **Criptografia:** Os dados em trânsito e em repouso são criptografados (TLS e AES-256), em total conformidade com a Lei Geral de Proteção de Dados (LGPD).
- **Trilha de Auditoria:** Cada decisão tomada ou campo preenchido pela IA gera um registro imutável nos logs do sistema, permitindo auditoria completa pelas equipes de governança do Senac.

5. Conclusão

A engenharia de software aplicada na stack proposta garante que o Senac SP receba uma solução de Inteligência Artificial de última geração, aderente a 100% dos requisitos do edital, mas com uma eficiência de consumo que viabiliza o modelo de licenciamento de preço fixo anual (parcela única) sem riscos operacionais ou surpresas financeiras para a instituição.