

Aquisição de serviços de consultoria na área da Inteligência Artificial B0069/2025

Ponto de situação

30/04/2026

PONTO DE SITUAÇÃO

Conquistas

- **Processamento e Ingestão:** Análise e benchmarking de ferramentas de processamento de documentos. Estudo de diferentes estratégias de chunking. Desenvolvimento de uma pipeline completa de processamento para ingestão de documentos na vectorstore.
- **Documentação:** Relatórios técnicos dos benchmarkings realizados (para as bases de dados e para a ferramenta de processamento de documentos).
- **Infraestrutura:** Centralizar os ficheiros da publicação dos componentes relevantes com a documentação associada aos mesmos.
- **UI:** Pesquisa e análise de opções de frontend disponíveis.

Em curso

- **Processamento e Ingestão:** Execução da pipeline de ingestão com os documentos definidos.
- **Documentação:** Relatório técnico do benchmarking realizado para o frontend.

Próximos passos

- **UI:** Desenvolvimento e publicação do frontend (versão base sem RAG), com integração na pipeline de publicação da infraestrutura completa.
- **Modelo Conversacional RAG:** Definição da pipeline conversacional. Desenvolvimento do modelo de classificação de intenções.

Riscos / Temas pendentes

- **Conectividade e Ambiente de Desenvolvimento:** A rede na DCI impõe restrições de proxy e interceção TLS que impedem a instalação de dependências Python acima de um determinado tamanho a partir do PyPI, inviabilizando o uso normal de ferramentas como pip.
  - ➢ Este tema foi ultrapassado com a instalação de um servidor PyPI privado no servidor interno, permitindo distribuir os pacotes Python sem depender de conectividade direta ao PyPI. As restrições de proxy da rede permanecem e recomendamos que sejam revistas.
- **Definição de UI:** Existe um frontend em Angular, desenvolvido especificamente para o Exército, que poderá ser reutilizado. Contudo, Angular está fora da especialidade da equipa atual, existindo já uma alternativa identificada em Python. Este tema requer alinhamento entre os stakeholders para definição do caminho a seguir.

ROADMAP

Dados de dia 30/04

Fase 0: Kick-off — 100%
60h

Data de Conclusão Estimada: 17/04

- Acesso à infraestrutura
- Criação do git
- Identificar origem dos documentos e selecionar amostra inicial
- Preparar conjunto Q&A para avaliação do modelo
- Disponibilização dos modelos on-prem

Fase 1: Ingestão Básica e Chunking — 100%
100h

Data de Conclusão Estimada: 17/04

- Pipeline com PDFs nativos
- Benchmarking de ferramentas de processamento dos documentos
- Definição da estratégia de chunking
- Definição de metadados a extrair

Fase 2: Vectorstore e Retrieval — 80%
100h

Data de Conclusão Estimada: 06/05

- Definição da base de dados vetorial
- Geração de embeddings e armazenamento dos documentos
- Configuração da vector search e retriever

Fase 3: Modelo Conversacional RAG
160h

Data de Conclusão Estimada: 03/06

- Preparação do modelo de classificação de intenções
- Construção da pipeline RAG com LLM
- Definição de métricas de avaliação
- Avaliação com o dataset preparado

Fase 4: Melhorias sobre Ingestão — 40%
80h

Data de Conclusão Estimada: 17/06

- Integração de OCR
- Extração de tabelas
- Sumarização de imagens e gráficos

Fase 5: Publicação — 30%
60h

Data de Conclusão Estimada: 26/06

- Definição do frontend
- Preparação da API / demonstração

axians

The best of ICT with a human touch

axians.com

VINCI ENERGIES