Desafio: IntelligentIA

Edital: CPSI

1. O Desafio (Solução Inovadora)

Uma plataforma de inteligência investigativa integra e cruza automaticamente dados de SIMBA, SITTEL e dispositivos móveis. Esta solução aplica Inteligência Artificial para identificar padrões, vínculos e cronologias, consolidando-os em um repositório institucional de conhecimento acumulativo com rastreabilidade total.

2. Objetivo

Plataforma de inteligência investigativa baseada em IA e arquitetura RAG (Retrieval-Augmented Generation), capaz de realizar a ingestão, indexação vetorial e análise correlacional de dados estruturados e não estruturados (SIMBA, SITTEL, extrações forenses e documentos).

A solução deve obrigatoriamente:

(i) permitir consultas em linguagem natural com respostas fundamentadas exclusivamente nos dados do caso (grounding);
(ii) apresentar interface conversacional para mídias sociais com áudios transcritos via ASR;
(iii) gerar grafos de relacionamento, com suporte a exportação em formatos abertos (como GraphML ou JSON-LD), e visualizações georreferenciadas automáticas; e
(iv) garantir validação humana estruturada (human-in-the-loop).

3. Oportunidades de Mercado

A solução possui alto potencial de escalabilidade, visando um mercado que abrange os 26 Ministérios Públicos Estaduais, o Ministério Público da União, além de Polícias Judiciárias e órgãos de Inteligência de Estado. A carência de ferramentas que integrem nativamente SIMBA e SITTEL com IA generativa cria um oceano azul para a contratada.

4. Requisitos Funcionais

ID	Descrição	PoC	Homologação
1	Gestão de Casos pelo usuário responsável, permitindo a segregação de dados por investigação e atribuição de acesso a outros usuários do sistema.	5	10
2	Carga de registros digitais por upload	5	15

ID	Descrição	PoC	Homologação
	manual com validação de formato e vinculação a Caso. Na fase de Homologação, incluir suporte a arquivos de grande volume (>2GB), upload retomável e processamento assíncrono com indicação de progresso.		
03	Ingestão de arquivos do SIMBA (layout BACEN TXT), com extração de campos estruturados e indexação para busca. Na Homologação, incluir suporte a extratos em PDF com extração via OCR/Parser de tabelas.	10	20
04	Ingestão dos registros do SITTEL (XML layout SITTEL), com parsing dos campos estruturados – origem, destino, data/hora, duração, tipo de comunicação e coordenadas de ERB – e indexação vetorial, segregados por Caso.	10	20
05	Ingestão de extrações forenses, com suporte inicial ao formato IPED (indexação de mensagens e contatos) na PoC. Na Homologação, expandir para formatos UFD/UFDX e parsing de artefatos de navegação.	5	20
06	Implementação de arquitetura RAG (Retrieval-Augmented Generation) sobre o acervo do caso, permitindo consultas em linguagem natural com respostas fundamentadas e	20	30

ID	Descrição	PoC	Homologação
	citação de fontes.		
07	Aplicação de filtros avançados, expressões regulares (REGEX), pesquisa por palavras-chave, operadores booleanos e consultas por similaridade semântica.	10	20
08	Implementação de mecanismos de Processamento de Linguagem Natural (PLN), com suporte a LLMs, para extração de entidades, identificação de padrões comunicacionais, classificação de conteúdos e geração de insights analíticos.	20	30
09	Cadastro de Operações e investigados, com vinculação estruturada dos vestígios a cada caso.	5	10
10	Integração com sistemas institucionais, bancos de dados relacionais (ODBC/JDBC), datalakes e bases externas via APIs (REST e SOAP), incluindo portais de transparência e dados abertos governamentais.	-	20
11	Geração de relatórios técnicos customizados em formatos abertos (como PDF, HTML ou DOCX), com inclusão automatizada de evidências de interesse, metadados relevantes e preservação da rastreabilidade.	5	15
12	Base de	5	15

ID	Descrição	PoC	Homologação
	conhecimento institucional compartilhada, gerenciada pelo Administrador, disponível como camada complementar ao RAG em todos os Casos: orientações técnico-jurídicas investigativas, tipologias criminosas e demais materiais.		
13	Visualização Georreferenciada de eventos investigativos (ERBs, locais de crime, endereços extraídos), com suporte a camadas de mapas e linha do tempo (timeline) integrada.	-	15
14	Processamento Multimodal (ASR e OCR): Transcrição automática de áudios/vídeos e extração de texto de imagens/PDFs escaneados para indexação no pipeline de RAG.	15	25
15	Visualização conversacional de extrações de dispositivos móveis (WhatsApp/Telegram), reconstruindo o chat com suporte a mídias, áudios transcritos e metadados de autoria.	10	20
16	Geração dinâmica de grafos de relacionamento entre alvos, pessoas, empresas, eventos e demais entidades correlatas, com visualização e navegação interativa (ex. Cytoscape) e integradas diretamente à	10	25

ID	Descrição	PoC	Homologação
	plataforma, permitindo exploração integral dos dados e registros processados. Os grafos deverão suportar correlação e representação unificada de todos os dados passíveis de ingestão pela aplicação, incluindo registros bancários, fiscais, telefônicos, telemáticos, extrações forenses de dispositivos, consultas a bases internas e externas e demais artefatos investigativos, possibilitando identificação de nós centrais, vínculos diretos e indiretos, padrões relacionais e fluxos de interação.		
17	Mecanismo de 'Human-in-the-loop' para validação de achados da IA, permitindo ao usuário confirmar ou descartar entidades e vínculos extraídos, retroalimentando a base do caso.	5	15

5. Requisitos Técnicos

ID	Item	Descrição	PoC	Homologação
01	-	Arquitetura Simplificada (MVP): Solução funcional demonstrando fluxos essenciais.	Sim	-
02	6.2.1	Segregação lógica: Isolamento absoluto de dados e interações.	Sim	Sim
03	6.2.2	Segurança da informação: Vedação de credenciais no código, comunicação via HTTPS com TLS	-	Sim

ID	Item	Descrição	PoC	Homologação
		1.2 ou superior, quando aplicável. Autenticação integrada à Microsoft Entra ID, por meio de protocolos padronizados de autenticação e autorização (ex.: OAuth 2.0 ou equivalentes).		
04	6.2.3	Autenticação Avançada: Configurável por perfil e obrigatória para administradores, quando houver interface administrativa, console de gestão ou operações privilegiadas. Deverá ser compatível com provedores de identidade institucionais.	-	Sim
05	6.2.4	Controle de Acesso (RBAC): Implementação de RBAC para segmentação granular e rastreabilidade.	-	Sim
06	6.2.5	Gestão de Segredos: Cofres (Vaults) com política de rotação de chaves. Aplicável quando a solução realizar armazenamento, processamento ou integração com credenciais, chaves, tokens de acesso ou segredos operacionais sensíveis. Em ambientes on-premise, admite-se o uso de mecanismos institucionais equivalentes do MPRS	-	Sim
07	6.2.6	Proteção de Rede: Defesa contra OWASP Top 10, quando aplicável à arquitetura proposta. Soluções com APIs, acesso	-	Sim

ID	Item	Descrição	PoC	Homologação
		externo ou serviços em nuvem deverão implementar rate limiting e mitigação contra DoS/DDoS. Soluções SaaS, cloud ou híbridas, a proteção de infraestrutura será responsabilidade da proponente.		
08	6.2.7	Arquitetura API First: Disponibilizando interfaces padronizadas e interoperáveis (ex.: REST, GraphQL, gRPC ou equivalentes), com autenticação segura, documentação técnica e versionamento controlado.	Sim	Sim
09	6.2.8	Conteinerização e Serverless: Imagens Docker/OCI com execução portável, permitindo serviços gerenciados/serverless e rollback.	-	Sim
10	6.2.9	Escalabilidade Horizontal: Quando aplicável à arquitetura da solução, balanceamento dinâmico para garantir escalabilidade.	-	Sim
11	6.2.10	Observabilidade: Quando aplicável à arquitetura da solução, emissão de métricas, logs e traços (OpenTelemetry ou equivalente).	-	Sim
12	6.2.11	FinOps (Gestão de Consumo) : Para itens com tarifação variável ou consumo elástico de recursos, limites/quotas parametrizáveis, alertas de consumo, rateio/atribuição e relatórios de auditoria financeira.	-	Sim

ID	Item	Descrição	PoC	Homologação
13	6.3.1	Trilhas de Auditoria Básicas: Garantia de integridade e rastreabilidade das operações.	-	Sim
14	6.3.2	Computação Forense (WORM): Armazenamento de logs em padrão WORM, integridade garantida por hash/criptografia e rastreabilidade da cadeia de custódia.	-	-
15	6.3.3	Inventário SBOM: Documentação SBOM (SPDX ou CycloneDX).	Sim	Sim
16	6.4.1	Prevenção a Injeção e Vazamento: Quando aplicável a soluções que utilizem modelos de Inteligência Artificial generativa ou LLMs, proteção contra manipulação (jailbreak) e vazamento de system prompts.	-	Sim
17	6.4.2	Mitigação de Alucinações (Grounded Generation): Condicionar respostas às fontes e evitar invenção de fatos/normativas.	Sim	Sim
18	6.4.2.1	Metodologia de Aferição Objetiva: durante a avaliação, a solução será submetida a um Dataset de Validação fornecido pelo MPRS.	-	Sim
19	6.4.2.2	Critério de Aceite: Quando aplicável a soluções que utilizem modelos de Inteligência Artificial generativa ou LLMs, o desempenho será aferido pela métrica HAM (Hallucination Rate), com HAM	-	Sim

ID	Item	Descrição	PoC	Homologação
		máximo de 10% sobre o conjunto de testes.		
20	6.4.3	Avaliação Contínua e Fine-Tuning: Quando aplicável a soluções baseadas em IA generativa ou RAG, uso de frameworks de qualidade técnica (como RAGAS) medindo fidelidade e consistência.	-	Sim
21	6.4.4	Trilha de Justificativa de IA: Rastreabilidade de ID, fontes/referências, parâmetros da execução e motivos de bloqueio.	Sim	Sim
22	6.4.6	Vedações Operacionais Mínimas: Bloqueio de respostas com "aparência de certeza" sem verificação, vedação de uso de dados reais não autorizados e bloqueio contra desativação de segurança por comandos.	-	Sim
23	6.4.6	Agnosticismo de Modelo de IA: Possibilidade de troca do LLM	-	Sim
24	6.4.7	Gestão de Prompts e Capacitação: Repositório versionado com controle de acesso e fornecimento de manuais/playbooks	-	Sim
25	6.4.8	Agentes Autônomos em Sandbox: Atuação de agentes autônomos em sandbox com limites configuráveis de autonomia e recursos, e bloqueio auditável de operações sensíveis ou escalonamento indevido de privilégios.	-	Sim
26	6.4.9	Métricas de Mídia: Aferição pela métrica	Sim	Sim

ID	Item	Descrição	PoC	Homologação
		WER (Word Error Rate), com WER máximo de 10% em conjunto de testes.		
27	6.2.12	Processamento Paralelo e Execução Assíncrona: Quando aplicável à solução, suportar processamento paralelo e execução assíncrona, evitando bloqueios sequenciais e degradação do fluxo principal por chamadas externas ou tarefas de longa duração.	-	Sim
28	6.4.10	Métricas de OCR: Aferição pela métrica CER (Character Error Rate), com CER máximo de 5% em conjunto de testes.	Sim	Sim