Desafio: Análise de Prova Digital

Edital: CPSI

1. O Desafio (Solução Inovadora)

Uma plataforma de análise forense digital integrada processa volumes massivos e heterogêneos de evidências, utilizando inteligência artificial para identificar padrões, correlações e vínculos. A solução oferece interfaces de visualização contextualizadas para comunicações móveis e mantém um registro auditável de todo o processo analítico.

2. Objetivo

Disponibilizar solução tecnológica destinada à análise integrada da prova digital, apta a realizar carga célere, processamento estruturado e correlação simultânea de múltiplos vestígios digitais, com rastreabilidade integral do processo analítico e conformidade com os requisitos de IA do MPRS.

A solução deverá permitir:

(i) ingestão segura e auditável de múltiplos formatos de extração forense (UFD, E01, SQLite etc.) com validação automática de hash;

(ii) processamento automatizado de dados estruturados e não estruturados;

(iii) transcrição automática de áudios/vídeos (ASR) com indexação;

(iv) OCR forense em documentos e imagens;

(v) indexação vetorial para busca semântica;

(vi) consultas em linguagem natural via LLM com arquitetura RAG, garantindo agnosticismo de modelo e métricas de controle de alucinação (HAM - Hallucination Rate);

(vii) extração automática de entidades e correlação de vestígios;

(viii) geração de grafos interativos de relacionamento;

(ix) filtros avançados e busca por similaridade;

(x) controle de integridade contínuo e logs em padrão WORM (imutáveis);

(xi) geração de relatórios técnicos auditáveis; e

(xii) visualização conversacional contextualizada de aplicativos de mensagens.

Busca-se reduzir o tempo médio de análise, elevar a qualidade técnica das investigações e garantir a sustentabilidade jurídica da prova digital diante de questionamentos sobre a metodologia analítica empregada, assegurando que a IA atue como assistente fundamentado e rastreável.

3. Oportunidades de Mercado

A análise de prova digital tornou-se o núcleo da persecução penal contemporânea, especialmente em crimes de corrupção e lavagem de dinheiro. O modelo operacional vigente, contudo, é limitado por análises compartimentadas e manuais, incapazes de tratar o conteúdo não estruturado em escala.

Oportunidade Institucional e Replicabilidade: A solução possui alto potencial de replicabilidade para outros ramos do Ministério Público, Polícias Judiciárias e laboratórios de perícia digital, todos enfrentando o mesmo desafio de volume de dados.

No modelo de CPSI, isso representa uma oportunidade estratégica para a contratada na exploração comercial da solução (Revenue Share), visando tornar-se o padrão tecnológico nacional para análise estruturada de prova digital.

Diferenciais de Inovação: A plataforma diferencia-se ao integrar quatro camadas estratégicas:

(1) Análise multimodal correlacionada;

(2) Visualização conversacional contextualizada;

(3) Camada analítica de grafos de relacionamento; e

(4) Camada cognitiva via IA com arquitetura RAG, que elimina alucinações e fornece respostas rastreáveis.

Tais diferenciais garantem não apenas eficiência operacional, mas segurança jurídica superior, transformando dados brutos em evidências robustas para o sistema de justiça criminal.

4. Requisitos Funcionais

ID	Descrição	PoC	Homologação
1	Gestão de Casos pelo Membro do MPRS com atribuição, possibilitando controle de acesso a outros usuários (Membros ou Servidores) ao respectivo Caso, rastreabilidade completa de todas as operações e, para homologação, suporte à importação de arquivos estruturados de Casos do sistema SIMBA ou integração via API.	5	15
2	Ingestão e processamento de extrações forenses de dispositivos móveis, com suporte a casos exportados/processados	15	25

ID	Descrição	PoC	Homologação
	pelo IPED e a extrações do ecossistema UFED (UFD/UFDX), incluindo parsing e indexação de mensagens, contatos e artefatos de navegação, com suporte a grandes volumes, upload retomável e processamento assíncrono.		
3	Transcrição automática de áudios e vídeos extraídos de dispositivos e aplicativos de mensagens, utilizando modelos ASR em português brasileiro, com identificação de locutores (diarização), sincronização temporal, indexação do conteúdo transcrito e exibição na interface conversacional.	10	20
4	Visualização conversacional de comunicações extraídas de aplicativos de mensagens (WhatsApp, Telegram e outros), com reprodução da sequência e contexto original das trocas, identificação dos participantes, exibição de áudios com transcrição vinculada em caixa de texto, possibilidade de acionar consultas RAG a partir de mensagens selecionadas como âncora de contexto, e rastreabilidade de cada mensagem ao vestígio de origem.	15	20
6	Processamento automatizado de imagens e PDFs digitalizados (OCR Forense) para extração de texto com suporte a múltiplos idiomas (foco	5	10

ID	Descrição	PoC	Homologação
	em PT-BR), integrando o conteúdo extraído ao índice de busca híbrida e semântica do Caso.		
7	Indexação vetorial do acervo do Caso, segmentada por operação investigativa, para busca semântica sobre dados textuais, transcritos e extraídos por OCR.	10	15
8	Busca híbrida e multimodal: capacidade de realizar buscas combinando operadores booleanos, expressões regulares (REGEX) e filtros por metadados técnicos com busca semântica vetorial sobre todo o acervo (textos, transcrições e OCR).	10	20
9	Extração automática de entidades (pessoas, locais, datas, organizações e valores) e identificação de padrões comunicacionais por Processamento de Linguagem Natural (PLN).	-	20
10	Consulta em linguagem natural ao acervo do Caso via LLM com arquitetura RAG (Chat Analítico), com respostas fundamentadas exclusivamente nos dados indexados, indicação precisa de fontes, histórico de sessões e controle de alucinação baseado na métrica HAM (Hallucination Rate) de até 10%.	20	30
11	Correlação automática entre diferentes tipos de vestígios do mesmo Caso (texto, áudio transcrito, imagem e documento) baseada	-	20

ID	Descrição	PoC	Homologação
	em entidades (pessoas, locais, organizações) e marcos temporais.		
12	Geração de grafos interativos de relacionamento para análise visual de vínculos entre investigados, contatos, eventos, dispositivos e transações, permitindo navegação dinâmica por nós e exportação do diagrama.	10	20
13	Automação de rotinas de conferência de integridade (re-hash) dos vestígios custodiados, com validação automática de hash (MD5/SHA-256) e alertas imediatos em caso de divergência.	5	10
14	Cadastro de operações e investigados com vinculação estruturada dos vestígios a cada Caso.	5	5
15	Disponibilização de APIs (REST) para integração com sistemas institucionais e suporte a conectores de dados, se aplicável.	-	20
16	Geração de relatórios técnicos analíticos em formatos exportáveis (PDF/HTML) contendo trechos selecionados, hashes originais dos vestígios, metadados de origem e trilha de auditoria da análise.	5	15
17	Registro detalhado de logs de acesso e operações, garantindo auditabilidade e, para a homologação, armazenamento em padrão WORM (Write Once, Read Many) para assegurar a imutabilidade da cadeia de custódia analítica.	5	15

ID	Descrição	PoC	Homologação
18	Repositório opcional de conhecimento forense (jurisprudência e manuais) indexado para busca semântica de apoio ao analista.	-	10
19	Carregamento de documentos e arquivos de interesse investigativo – relatórios, registros de SITTEL, dados de outras fontes e demais materiais relevantes – pelo Membro do MPRS com atribuição ou servidor autorizado, com registro de metadados de origem, integração ao acervo vetorial do Caso, disponibilidade para consulta via RAG e correlação automática com os vestígios forenses do Caso.	15	15
20	Painel de Casos e Status de Processamento: A tela inicial do usuário autenticado deve exibir os Casos aos quais tem acesso, com indicadores visuais do status de cada operação de ingestão (em fila, em processamento, concluída, com erro), percentual de conclusão e alertas de falha, sem necessidade de acesso a módulos internos para verificar o andamento.	5	15
21	Fluxo Guiado de Criação de Caso e Ingestão: O sistema deve conduzir o usuário por assistente passo a passo para criação de novo Caso e carga de vestígios, com validação automática de formato, exibição do hash calculado na ingestão,	10	10

ID	Descrição	PoC	Homologação
	confirmação de integridade e feedback visual de cada etapa, tornando o processo acessível a usuários sem perfil técnico avançado.		
22	Linha do Tempo Unificada do Caso: O sistema deve gerar automaticamente uma linha do tempo cronológica consolidando eventos de diferentes vestígios do mesmo Caso — mensagens, chamadas, transações, documentos e registros de localização —, com filtros por tipo de evento, participante e período, e correlação visual entre eventos simultâneos ou próximos.	10	20
23	Comparação entre Dispositivos ou Períodos: O analista deve poder selecionar dois dispositivos, dois participantes ou dois períodos e visualizar, em painel comparativo lado a lado, as comunicações, entidades extraídas e eventos correspondentes, com destaque automático para coincidências e divergências relevantes.	-	15
24	Exportação de Linha do Tempo Interativa: O sistema deve permitir exportar a linha do tempo do Caso em formato HTML autônomo ou PDF interativo, com navegação por eventos, adequado para apresentação em audiências ou anexação em peças, contendo os	5	10

ID	Descrição	PoC	Homologação
	hashes dos vestígios de origem de cada evento exibido.		

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 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).	-	Sim
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:	-	Sim

ID	Item	Descrição	PoC	Homologação
		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		
07	6.2.6	Proteção de Rede: Defesa contra OWASP Top 10, quando aplicável à arquitetura proposta. Soluções com APIs, acesso 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.	-	Sim
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	-	Sim

ID	Item	Descrição	PoC	Homologação
		da solução, balanceamento dinâmico para garantir escalabilidade.		
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
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.	-	Sim
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	Sim	Sim

ID	Item	Descrição	PoC	Homologação
		Generation): Condicionar respostas às fontes e evitar invenção de fatos/normativas.		
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	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 máximo de 10% sobre o conjunto de testes.	-	Sim
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	-	Sim

ID	Item	Descrição	PoC	Homologação
		Modelo de IA: Possibilidade de troca do LLM.		
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 WER (Word Error Rate), com WER máximo de 10% em conjunto de testes.	Sim	Sim
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