Desafio: Anonimização

Edital: CPSI

1. O Desafio (Solução Inovadora)

A solução consiste em um sistema tecnológico para identificação e pseudonimização padronizada de dados pessoais em documentos institucionais. Uma camada proxy, integrada ao portal, processará dinamicamente a pseudonimização de visualizações e arquivos, conforme o nível de acesso do usuário, preservando a versão original do documento.

2. Objetivo

O presente desafio tem por objetivo fomentar o desenvolvimento de solução tecnológica integrada e robusta de pseudonimização de documentos, que vá além do simples tarjamento ou ocultação manual de dados pessoais ostensivos em documentos tratados nos sistemas do Ministério Público, que venham a ser compartilhados com terceiros ou divulgados nos portais do MPRS.

Busca-se solução capaz de atuar de forma transversal ao ciclo de produção documental, abrangendo tanto a elaboração de peças institucionais e relatórios internos quanto o fornecimento de cópias de expedientes e a disponibilização de documentos ao público externo, assegurando a proteção de dados pessoais e informações sensíveis sem prejuízo à transparência e à inteligibilidade dos atos institucionais.

Nesse sentido, a solução deverá ser tecnicamente apta a analisar documentos estruturados e não estruturados, identificar dados pessoais e dados sensíveis inseridos em linguagem jurídica contextualizada, e aplicar decisões de pseudonimização com base em regras legais, administrativas e de negócio, alinhadas à LGPD, à Lei de Acesso à Informação (LAI) e às legislações do CNJ.

Nos últimos anos, o Brasil foi palco de sucessão de incidentes graves de segurança da informação envolvendo tanto órgãos públicos quanto instituições financeiras, evidenciando a fragilidade estrutural na proteção de dados pessoais. Entre os casos mais emblemáticos estão o vazamento de dados sensíveis de pessoas vivendo com HIV em Feira de Santana (BA), decorrente da publicação indevida de ato administrativo; o vazamento de senhas de acesso a sistemas internos do Tribunal de Contas da União (TCU); a exposição de chaves Pix e informações cadastrais de clientes da Caixa Econômica Federal; além do chamado “megavazamento” de 2021, que expôs dados de aproximadamente 223 milhões de brasileiros, incluindo informações de pessoas falecidas. Soma-se a esse cenário a falha em sistemas do Ministério da Saúde que resultou na exposição de dados pessoais de mais de 200 milhões de cidadãos, envolvendo informações altamente sensíveis relacionadas à saúde.

Assim, a implementação de camada intermediária de acesso (proxy), integrada aos sistemas e ao portal institucional, constituirá mecanismo técnico relevante para mitigar e até prevenir incidentes de segurança e exposições indevidas de dados pessoais como os recentemente observados no setor público e financeiro. Ao permitir que o documento permaneça armazenado

em sua forma original nos sistemas internos, enquanto a visualização e a geração de cópias são realizadas de forma dinâmica e pseudonimizada conforme o nível de acesso do usuário, o proxy reduz significativamente o risco de publicações equivocadas, acessos excessivos ou compartilhamentos não autorizados. A inexistência de padrão “universal” de pseudonimização, substituído por decisões caso a caso, orientadas por regras legais e perfis de acesso, possibilitará maior aderência ao princípio da necessidade e da minimização de dados, dificultando a exposição massiva e indiscriminada de informações sensíveis, como aquelas verificadas em vazamentos envolvendo dados de saúde, credenciais de sistemas internos e bases cadastrais de grande escala.

A ferramenta deve, ainda, preservar a compreensão e o valor informativo dos documentos resultantes, garantindo que o conteúdo permaneça inteligível ao seu destinatário legítimo, ao mesmo tempo em que assegure rastreabilidade, auditabilidade e integridade de todas as operações realizadas, de modo a permitir controle institucional, verificação posterior das decisões adotadas e mitigação de riscos jurídicos e operacionais.

3. Oportunidades de Mercado

Sistema de Justiça Público (G2G): A necessidade de pseudonimização e proteção de dados em documentos oficiais não é exclusiva do MPRS, sendo estrutural a todo o sistema de justiça brasileiro, incluindo Ministérios Públicos Estaduais, Ministério Público da União, Defensorias Públicas, Advocacias Públicas, Controladorias e Órgãos de Controle que enfrentam desafios idênticos diante da LGPD e do impositivo dever de transparência. A solução possui alto potencial de replicabilidade, especialmente considerando que servidor Proxy funcionará como filtro de segurança integrado a sistemas existentes de processo eletrônico, portais de transparência e gestão documental.

Vertical de Justiça (Tribunais): Tribunais de Justiça, Tribunais Regionais e Cortes Superiores produzem e publicam diariamente grande volume de decisões, acórdãos, despachos e documentos judiciais que contêm dados pessoais de partes, testemunhas e terceiros. A pressão regulatória exercida pelo CNJ para ampliação da transparência, aliada às exigências da LGPD, cria um ambiente propício para soluções que automatizem a pseudonimização de documentos judiciais, assegurando publicidade responsável e mitigação de riscos legais. Plataforma capaz de integrar IA aplicada à análise jurídica e proteção de dados representa ativo estratégico para o Judiciário.

Administração Pública e Órgãos de Controle: Secretarias, autarquias, fundações públicas, agências reguladoras e tribunais de contas enfrentam, de forma recorrente, o desafio de conciliar o dever de transparência ativa com a proteção de dados pessoais na divulgação de informações por meio de portais públicos, relatórios, pareceres e processos administrativos. Esse dilema se manifesta especialmente na necessidade de dar publicidade aos atos administrativos sem expor, de forma indevida, dados pessoais ou sensíveis dos titulares envolvidos. Ademais, Estados, Municípios e demais entes federativos, com destaque para órgãos das áreas de saúde e educação, lidam cotidianamente com grandes volumes de informações pessoais, muitas delas sensíveis, estando igualmente sujeitos às exigências legais de transparência e às obrigações de proteção de dados pessoais, o que reforça a necessidade de soluções técnicas estruturadas que assegurem conformidade normativa, segurança da informação e respeito aos direitos fundamentais dos cidadãos. A adoção de solução padronizada de pseudonimização pode se

tornar instrumento transversal de governança de dados, com escalabilidade nacional, especialmente em iniciativas coordenadas de governo digital.

Setor Privado Estratégico (B2B): No setor privado, escritórios de advocacia, departamentos jurídicos corporativos, empresas de compliance, auditoria e ESG, além de LegalTechs, também lidam com altos volumes de documentos que exigem compartilhamento externo sem exposição indevida de dados. Nesse contexto, ao estruturar este desafio, o MPRS busca não apenas solucionar demanda interna, mas fomentar o desenvolvimento de tecnologia replicável, com potencial de se tornar referência nacional em pseudonimização documental no setor público e privado, assegurando a ampliação da proteção dos dados pessoais dos titulares.

4. Requisitos Funcionais

ID	Descrição	PoC	Homologação
01	Identificar dados pessoais como: nome, data de nascimento, número de documento de identidade (CPF, RG e CNH), endereço residencial, telefone/celular, e-mail.	Sim	Sim
02	Identificar dados pessoais como: estado civil, filiação e nacionalidade relacionados a uma pessoa física.	20	25
03	Identificar dados pessoais como: placa de veículo, IP computador, IMEI celular.	25	Sim
04	Identificar dados pessoais como: número de documento de identificação (PIS/PASEP, carteira de trabalho, título eleitoral, matrícula funcional), dados profissionais (cargo, local de trabalho).	20	30
05	Identificar dados pessoais como: assinatura manuscrita e digital.	-	40
06	Identificar dados pessoais sensíveis como: origem racial ou étnica, convicção religiosa e opinião política, relacionados a	30	45

ID	Descrição	PoC	Homologação
	uma pessoa física.		
07	Identificar dados pessoais sensíveis como: filiação a sindicato ou a organização de caráter religioso, filosófico ou político, relacionados a uma pessoa física.	20	30
08	Identificar dados pessoais sensíveis como: dado referente à saúde.	40	Sim
09	Identificar dados pessoais sensíveis como: dado referente à vida sexual (orientação sexual, identidade de gênero, etc) relacionados a uma pessoa física.	25	30
10	Identificar dados pessoais de crianças e adolescentes, sinalizando que tais dados se referem a pessoa menor de idade.	45	Sim
11	Identificar dados de pessoas naturais, informações identificadoras de pessoas jurídicas e de figuras notórias (Ministros, Magistrados, Membros do Ministério Público, etc.), permitindo a aplicação de regras distintas de proteção ou transparência em consonância com as atribuições institucionais.	25	25
12	Identificar e distinguir dados pessoais de sujeitos homônimos (ex: 'João Silva, o réu' vs 'João Silva, o juiz') por meio de análise contextual	15	30
13	Identificar dados pessoais sensíveis: fotografias e imagens de humanos e	20	Sim

ID	Descrição	PoC	Homologação
	tatuagens relacionados a uma pessoa física.		
14	Disponibilizar etapa intermediária entre a identificação dos dados e a pseudonimização, permitindo a revisão manual, edição do documento e escolha da técnica de pseudonimização a ser aplicada.	-	Sim
15	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 01.	Sim	Sim
16	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 02.	20	Sim
17	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 03.	25	Sim
18	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 04.	20	30
19	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 05.	-	40
20	Pseudonimizar (hachurar,	30	45

ID	Descrição	PoC	Homologação
	mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 06.		
21	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 07.	20	30
22	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 08.	40	Sim
23	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 09.	25	30
24	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 10.	Sim	Sim
25	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 11.	25	25
26	Pseudonimizar (hachurar, mascarar/embaralhar, substituir, etc) dados pessoais em documentos identificados no requisito 12.	15	30

ID	Descrição	PoC	Homologação
27	Pseudonimizar (hachurar/tarjar) dados pessoais em documentos identificados no requisito 13.	20	Sim
28	Suportar a leitura e processamento de documentos nos formatos DOCX, PDF e conteúdo em texto puro (plain text) via integração.	20	Sim
29	Gerar documentos de saída em formato PDF sanitizado, garantindo que os dados pseudonimizados sejam removidos de forma irreversível das camadas de texto e metadados.	10	Sim
30	Realizar pseudonimização reversível (tokenização ou identificadores substitutos (asterisco, traço, iniciais, etc), mantendo correlação interna segura sob controle exclusivo do MPRS.	10	Sim
31	Garantir que um mesmo dado pessoal identificado (ex: 'João da Silva') seja substituído pelo mesmo token (ex: '[PESSOA_1]') em todo o documento e em documentos relacionados na mesma sessão, preservando a inteligibilidade do texto.	15	Sim
32	Implementar a interceptação de tráfego via Proxy ou camada de middleware, demonstrando a capacidade de processar e modificar o conteúdo de documentos em tempo real antes da entrega ao cliente final.	20	Sim

ID	Descrição	PoC	Homologação
33	Aplicação de Proxy com necessidade de alterações mínimas nos sistemas de origem.	-	40
34	Gerir perfis de acesso com diferentes níveis de permissão, assegurando controle de visualização, edição, validação e reversão da pseudonimização.	-	Sim
35	Registrar de forma automática e auditável as informações técnicas relacionadas ao acesso e uso da solução, incluindo dispositivo, endereço IP, localização aproximada, data e hora das interações.	-	Sim
36	Disponibilizar trilha de auditoria funcional que permita visualizar a regra de negócio ou o parâmetro legal que fundamentou cada decisão de pseudonimização realizada pela solução.	15	Sim
37	Manter histórico de versões das políticas e regras de pseudonimização aplicadas, em atendimento a normas e regulamentos, permitindo a reprodução exata de como um documento foi visualizado em determinada data/hora.	-	20
38	Capacidade de explicar, de forma compreensível e auditável, os critérios utilizados para identificar e pseudonimizar dados pessoais.	-	Sim
39	Mecanismo de certificação de integridade (como geração de hash ou equivalentes, com data	20	Sim

ID	Descrição	PoC	Homologação
	de emissão) que impeça adulterações posteriores e assegure a confiabilidade no documento original e no gerado.		
40	Disponibilização de interfaces intuitivas, com foco na experiência do usuário interno e externo, reduzindo a complexidade operacional da pseudonimização.	-	60
41	Geração de score de atendimento/adequação da pseudonimização realizada nos documentos.	-	20
42	Gerar relatórios gerenciais sobre volume de documentos processados, tipos de dados identificados, pseudonimizações realizadas.	-	20
43	Disponibilizar visualização gráfica ou estruturada de todos os dados pessoais identificados em um documento, facilitando a compreensão e a validação da pseudonimização.	20	20

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
03	6.2.2	Segurança da informação: Vedação de credenciais no código, comunicação	-	Sim

ID	Item	Descrição	PoC	Homologação
		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).		
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	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	-	Sim

ID	Item	Descrição	PoC	Homologação
		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.		
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	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.	Se aplicável	Se aplicável
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	-	-

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.	-	-
28	6.4.10	Métricas de OCR: Aferição pela métrica CER (Character Error Rate), com CER máximo de 5% em conjunto de testes.	-	-