

Desafio: Coletiv.ia

Edital: CPSI

1. O Desafio (Solução Inovadora)

Uma plataforma de e-discovery fático automatizada processará reclamações individuais em PDF, empregando Processamento de Linguagem Natural (PLN) para extrair a essência do problema. O sistema correlacionará esses dados com padrões de conduta sistêmicos e casos análogos da internet aberta, identificando proativamente lesões a direitos coletivos e homogêneos e gerando insights acionáveis.

2. Objetivo

Objetivo deste desafio é o desenvolvimento do Coletiv.ia, solução de inteligência analítica que atua como motor de busca semântica para caracterizar o caráter difuso de queixas de consumo. A solução deve permitir que o usuário suba PDF ou informe o nome de empresa para que a IA extraia o núcleo fático da reclamação e identifique padrões de conduta similares em fontes abertas.

Resultados Esperados:

Identificação de Padrões Fáticos: Superar buscas por palavras-chave, detectando similaridades semânticas que comprovem que o problema é comum e atinge a coletividade.

Abrangência Setorial: Capacidade de apontar quando lesão não é exclusiva de empresa, mas prática abusiva disseminada em setor econômico.

Fortalecimento Probatório: Geração de dossiês analíticos com prints certificados (Hash), volumetria e recorrência, servindo de evidência robusta para a instrução de Inquéritos Cíveis. Integração Institucional: Diálogo com o sistema SIM para cruzar achados externos com o histórico investigativo do MPRS.

3. Oportunidades de Mercado

A Coletiv.ia possui um Market-Fit universal no ecossistema de GovTech e LegalTech:

Sistema de Justiça Público: Replicabilidade imediata em todos os Ministérios Públicos (MPU e Estados) e Defensorias Públicas do Brasil.

Vertical de Justiça: Tribunais de Justiça e o CNJ para a gestão de Incidentes de Resolução de Demandas Repetitivas (IRDR).

Administração Pública: PROCONs, agências reguladoras (ANATEL, ANEEL) e Tribunais de Contas que necessitam fiscalizar condutas de mercado e conformidade regulatória.

Setor Privado: Departamentos de Compliance de grandes empresas e bancas de advocacia corporativa (Big Law) interessadas na prevenção de passivos judiciais em massa.

4. Requisitos Funcionais

ID	Descrição	PoC	Homologação
01	Ingestão de arquivos PDF (nativos ou digitalizados) com extração automática da essência fática. O suporte a OCR é obrigatório para a fase de homologação.	15	Sim
02	Busca semântica de similaridade em fontes abertas de defesa do consumidor (Reclame Aqui, Consumidor.Gov e outros), utilizando técnicas de web scraping, integração via API ou simulação por base de dados local na fase de PoC, com base no contexto fático extraído.	Sim	Sim
03	Identificação de múltiplos problemas distintos (concurso de lesões) dentro de um mesmo relato.	25	50
04	Deteção de similaridade de conduta entre diferentes empresas do mesmo segmento de mercado.	20	45
05	Funcionalidade proativa: monitoramento automático de temas sensíveis pré-definidos pelo MPRS.	5	20

ID	Descrição	PoC	Homologação
06	Filtro de segmentação geográfica para delimitar a análise ao estado do Rio Grande do Sul, respeitando a atribuição territorial do MPRS.	5	10
07	Anonimização ou mascaramento de dados pessoais sensíveis dos reclamantes (LGPD) antes do processamento final, flexibilizada a sua exigência na fase de PoC para priorizar a validação técnica.	-	Sim
08	Emissão de alerta de 'Anomalia de Mercado' ao detectar desvios estatísticos (picos) superiores a 2 desvios-padrão na média de queixas similares.	15	30
09	Funcionalidade 'Snapshot': captura de conteúdo da URL da prova com geração de Hash (SHA-256) e registro de data/hora.	10	30
10	Comparativo fático: confrontar a reincidência da empresa X com a média das concorrentes.	10	25
11	OCR avançado para extrair dados de prints de conversas ou fotos anexadas aos documentos.	25	50
12	Identificação de 'Zonas de Baixa Notificação' através do cruzamento georreferenciado entre volume de	5	10

ID	Descrição	PoC	Homologação
	queixas e densidade populacional/IDH da região.		
13	Interface para seleção de "Grau de Similaridade Semântica" (ajuste de sensibilidade da IA).	15	15
14	Exportação de visualizações de dados (grafos de relation entre empresas e linhas do tempo de reclamações) em formatos editáveis ou de alta resolução (SVG/PDF).	5	10
15	Recusa explícita e fundamentada quando a base de dados não sustentar a similaridade (Grounding).	Sim	Sim
16	Interface de Curadoria e Feedback (Human-in-the-loop): Permitir que o usuário valide ou corrija a classificação semântica feita pela IA, utilizando esses dados para o refinamento contínuo dos resultados da busca.	5	10
17	O sistema deve agrupar automaticamente as reclamações similares em 'clusters' temáticos, gerando um resumo automático (sumarização) do dano coletivo identificado em cada grupo.	20	30
18	Motor de Busca Semântica com Embeddings: Capacidade de	Sim	Sim

ID	Descrição	PoC	Homologação
	identificar similaridade entre queixas usando linguagem natural, superando a busca por palavras-chave (ex: identificar que 'cobrança indevida' e 'valor não reconhecido na fatura' referem-se ao mesmo núcleo fático).		
19	Extração de Entidades Estruturadas (NER): A solução deve extrair automaticamente CNPJ, nome da empresa e produto/serviço citado. A extração do 'valor do dano' deve ser tratada como campo opcional, condicionada à presença explícita e inequívoca no texto original.	10	15
20	Geração de Dossiê Consolidado: Funcionalidade para gerar relatório PDF contendo resumo fático, volumetria de casos similares, análise de impacto setorial e evidências (links/prints) acompanhadas de seus respectivos hashes (SHA-256) para garantia de integridade.	15	30
21	Integração com Ecossistema MPRS (SIM): Capacidade de consumir APIs ou serviços disponibilizados pelo MPRS para verificar se o núcleo fático (entidade/fato) identificado pela IA já	-	20

ID	Descrição	PoC	Homologação
	possui procedimentos investigativos instaurados, exibindo o status e número do protocolo na interface da solução.		
22	Deduplicação Inteligente de Queixas: Mecanismo para identificar e agrupar reclamações idênticas postadas pelo mesmo usuário em múltiplas plataformas, evitando a inflação artificial da volumetria do dano.	15	40
23	Análise de Recorrência e Tendência Temporal: Identificação automática se o padrão fático detectado apresenta crescimento súbito (surto de queixas) ou sazonalidade, com geração de gráfico de linha do tempo.	10	25

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	-	Sim

ID	Item	Descrição	PoC	Homologação
		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 externo ou serviços em nuvem deverão implementar rate limiting e mitigação contra DoS/DDoS.	-	Sim

ID	Item	Descrição	PoC	Homologação
		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	Containerização e Serverless: Imagens Docker/OCI com execução portátil, 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
13	6.3.1	Trilhas de Auditoria Básicas: Garantia de integridade e rastreabilidade das operações.	-	Sim

ID	Item	Descrição	PoC	Homologação
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 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	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	-	Sim

ID	Item	Descrição	PoC	Homologação
		baseadas em IA generativa ou RAG, uso de frameworks de qualidade técnica (como RAGAS) medindo fidelidade e consistência.		
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 WER (Word Error Rate), com WER máximo de 10% em conjunto de testes.	-	Sim
27	6.2.12	Processamento	-	Sim

ID	Item	Descrição	PoC	Homologação
		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.	Sim	Sim