

RELATÓRIO TRIMESTRAL DE MONITORAMENTO DE IA

Arts. 41 e 42 da Resolução CNJ nº 615/2025

Documento:	Relatório Trimestral de Monitoramento de Soluções de IA
Código:	IA-07 (Pós-Produção / Sustentação)
Base normativa:	Arts. 41, 42, 18 e 25 da Resolução CNJ nº 615/2025
Período de referência:	Julho a Setembro de 2026
Processo SEI:	0007600-34.2024.8.01.0000
Responsável:	SETIC/SUDIG/DIDIA
Versão / Data:	1.0 — 01/10/2026

HISTÓRICO DE REVISÕES

Versão	Data	Autor	Descrição
1.0	01/10/2026	Gabriel Guedes	Versão inicial

0. Orientações Gerais

Este relatório consolida o monitoramento contínuo das soluções de IA em produção no TJAC (Art. 41), com periodicidade mínima anual. Pode ser elaborado por solução ou de forma consolidada pela Divisão de IA e RPA. As informações alimentam a publicação no catálogo do CNJ (Art. 25) e dialogam com o relatório anual do Comitê Nacional (Art. 18).

1. Identificação e Visão Geral do Projeto

Nome	ADA – Assistente Digital Ampliada
Tribunal	TJAC SETIC/Divisão IA e RPA
Situação	Em produção
Classificação risco	Baixo (Art. 11 §3º CNJ 615)

2. Arquitetura Técnica e Modelos (LLMOps)

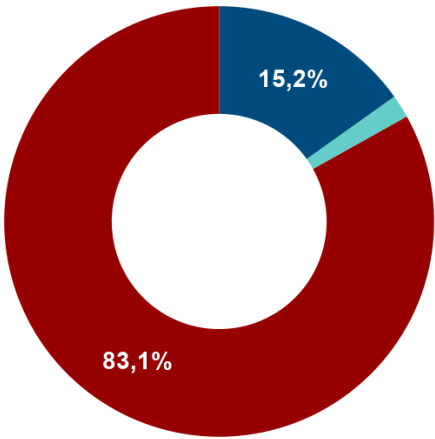
No período, a solução operou com três modelos de linguagem, com predominância do Gemini 2.5 Pro. A tabela abaixo apresenta os modelos utilizados e as camadas de mitigação de alucinação empregadas, que atuam de forma combinada: a base RAG fornece contexto documental confiável, os embeddings garantem busca por similaridade semântica, a engenharia de prompt orienta o comportamento do modelo, e a validação humana (Arts. 32 e 34 da Resolução CNJ nº 615/2025) confere a etapa final de checagem na entrega ao usuário.

Modelos	Fornecedor / Hospedagem	Uso no Período
gemini-2.5-pro	GOOGLE / Serpro	16.186
gemini-3.1-preview	GOOGLE / Serpro	1.860
gpt-4.1	Servidor	88.724
RAGs		
Base PANGEA		
Embeddings		
Vetorização Semântica para busca		
Mitigação Alucinação		
RAGs		
Engenharia de Prompt		
Validação Humana (Art. 32, 34 CNJ 615)		
Filtro de Segurança		

O gráfico mostra a proporção de requisições atendidas por cada modelo no trimestre. O Gemini 2.5 Pro concentrou a maior parte do uso.

Uso por LLM

- gemini-2.5-pro
- gemini-3.1-preview
- gpt-4.1



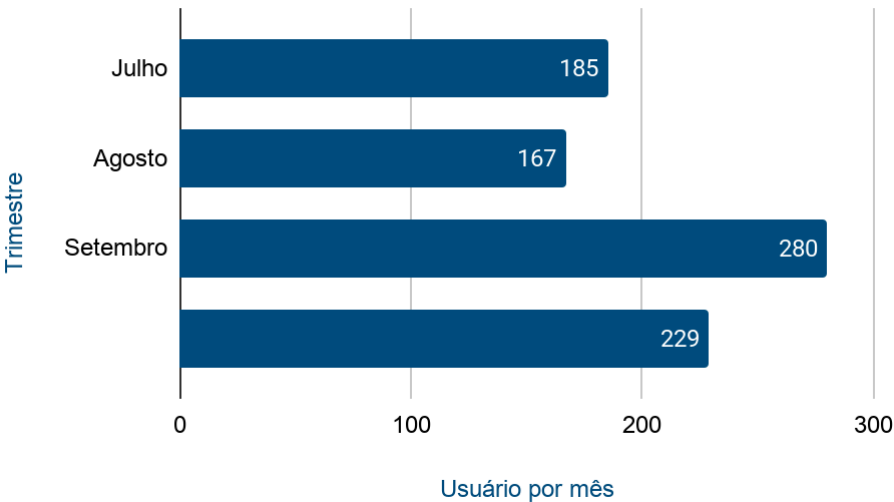
3. Métricas de Usabilidade

A tabela resume o uso da solução no trimestre de julho a setembro de 2026. Os gráficos detalham a evolução mês a mês.

Métrica	Julho a Setembro de 2026
Usuários Ativos	229
Qtd. de Requisições	244.199
Consulta de Precedentes	4.791
Total Transcrições	4.033
Total de Tokens	2.022.128.745

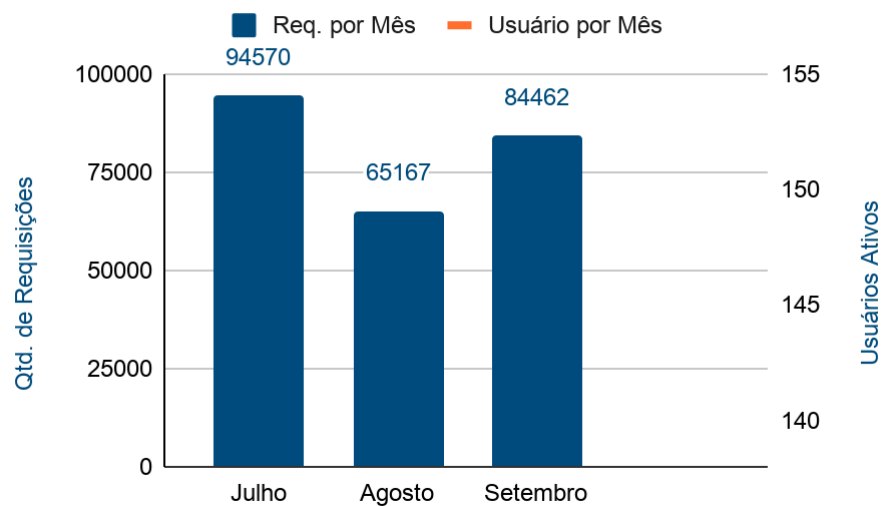
O número de usuários ativos caiu de 185 em julho para 167 em agosto e chegou a 280 em setembro, o maior valor do trimestre (+51% em relação a julho).

Usuário por mês no Período



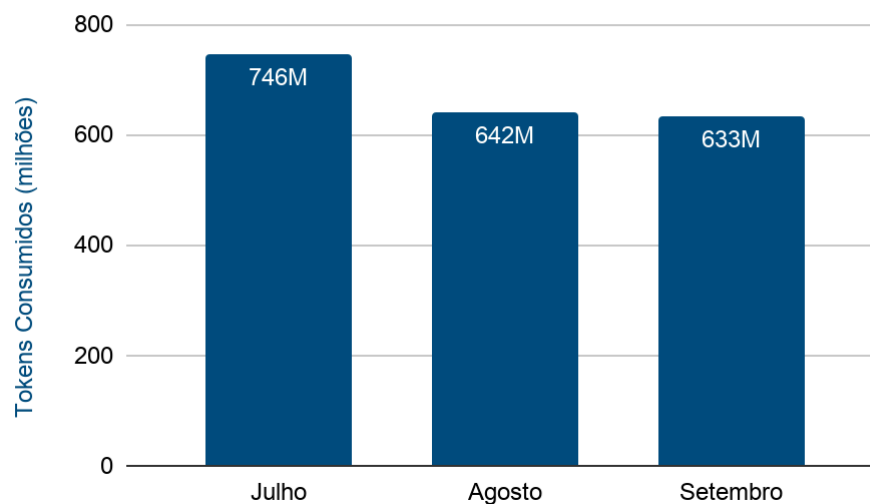
As requisições somaram 94.570 em julho, 65.167 em agosto (-31%) e 84.462 em setembro (+30% sobre agosto).

Evolução Mensal de Uso - Trimestre



O consumo de tokens foi de 746 milhões em julho, 642 milhões em agosto e 633 milhões em setembro, queda de cerca de 15% no trimestre.

Consumo de Tokens por Mês



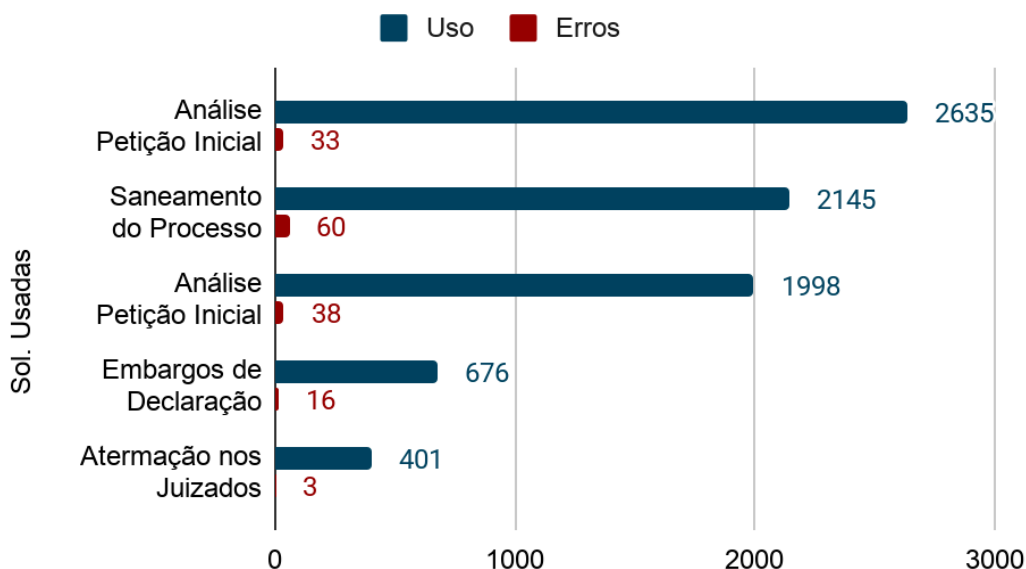
4. Desempenho Técnico

A tabela apresenta o volume de requisições por serviço (os cinco mais utilizados no período) e a quantidade de erros técnicos registrados. Os erros correspondem a falhas de comunicação entre a aplicação e os servidores dos modelos (erros HTTP 500), geralmente associadas a *timeout* e não a erros de conteúdo ou de qualidade das respostas.

Soluções	Requisições	Erros	Taxa de Erro (%)
Análise Petição Inicial Genérica	2.635	33	1,25%
Saneamento do Processo	2.145	60	2,80%
Análise Petição Inicial FIRAC+ (Prof. George Marmelstein)	1.998	38	1,90%
Embargos de Declaração Cível 1º Grau	676	16	2,37%
Atermação nos Juizados Especiais Cíveis	401	3	0,75%

Comparativo entre volume de uso e quantidade de erros técnicos (500/*timeout*) por solução. A leitura conjunta ajuda a identificar quais soluções são mais sensíveis a instabilidades de conexão com os servidores dos modelos.

Desempenho Técnico por Solução



5. Segurança Ofensiva: Prompt Injection & Defesas

No trimestre (julho a setembro), foram realizadas 88.368 checagens de prompt injection. O volume cresceu mês a mês, de 25.708 em julho para 34.943 em setembro (aumento de 36%).

Nesse período, 176 documentos foram sinalizados e classificados como falsos positivos após revisão do usuário (0,20% do total). Não houve tentativas de prompt injection confirmadas.

Trimestre	Qtd. Prompt Check	Falso Positivo
Julho	25708	42
Agosto	27717	69
Setembro	34943	65

A taxa de falsos positivos subiu em agosto, mas voltou a cair em setembro, mesmo com o maior volume do período. Em comparação com o relatório anterior (junho, 0,6%) a taxa atual é significativamente menor, e a funcionalidade segue estável, sem impacto relevante para os usuários.

6. Anonimizador: Desempenho e Conformidade

O módulo de anonimização segue ativo e em funcionamento em produção, mascarando dados pessoais e sensíveis antes do envio aos modelos de linguagem.

A contabilização do volume processado depende do envio dos logs ao Graylog. Essa funcionalidade já está desenvolvida e validada em homologação, aguardando apenas a publicação em produção. Por isso, não há dados quantitativos neste trimestre. Os logs do período estão preservados no servidor. Após a publicação, as métricas passarão a ser apresentadas nos próximos relatórios.

7. Eventos Adversos e Gestão de Riscos (Art. 42 CNJ 615)

Não foram registrados eventos adversos no período de julho a setembro de 2026. A tabela abaixo será preenchida conforme ocorrências futuras, com data, descrição, causa e medidas adotadas.

Data	Descrição do Evento	Causa	Medidas Adotadas