

# Projeto CITO

## Monitoramento e Análise de Jurisprudência do STF

### 1. Definição da problemática

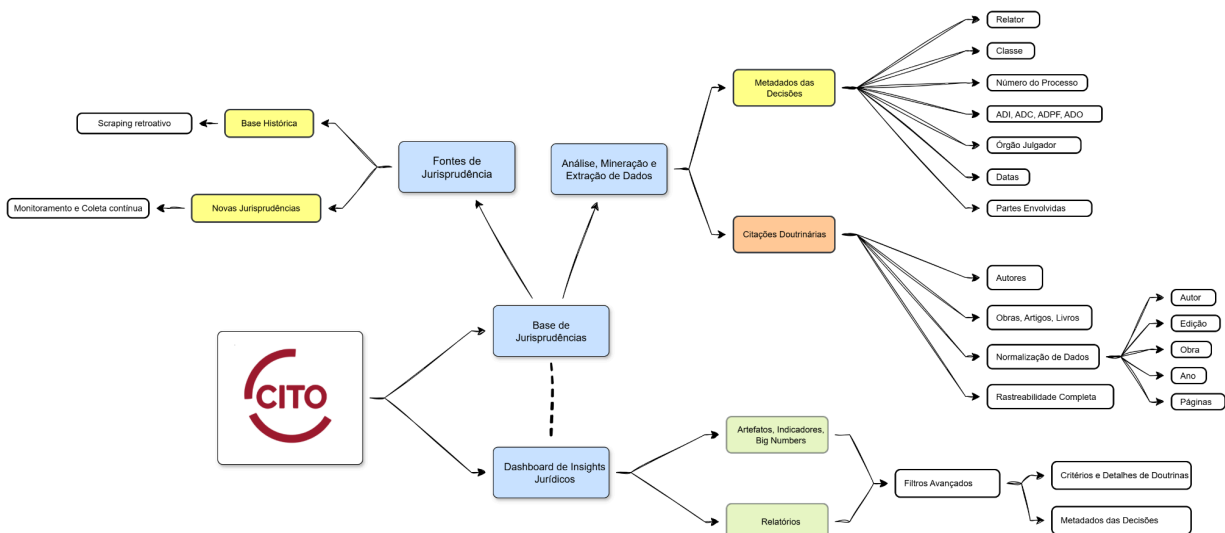
Plataforma para monitorar, coletar, estruturar e analisar decisões do STF, com foco em metadados jurídicos e citações doutrinárias. Oferece busca avançada, dashboards interativos e rastreabilidade documental, permitindo análises rápidas, confiáveis e estruturadas.

### 2. Dores

- **Volume elevado** e **dispersão** de publicações.
- **Dificuldade de acesso** ao inteiro teor e anexos.
- **Ausência de padronização** de metadados.
- **Complexidade** na identificação de citações.
- **Carência** de dashboards e relatórios analíticos.

### 3. Proposta

Solução integrada de **monitoramento** contínuo, **processamento** documental e **extração de dados**, transformando decisões em **conteúdo pesquisável e analítico**, útil para juristas, pesquisadores e instituições.



---

## 4. Objetivos e Entregas de Valor

- Automatizar a **coleta e preservação** de decisões do STF.
  - Estruturar **metadados jurídicos padronizados** (processo, classe, relator, órgão, datas, partes).
  - Extrair e padronizar **citações doutrinárias** (autores, obras, artigos).
  - Disponibilizar **busca full-text**, filtros avançados e consultas simples por processo e relator.
  - Fornecer **dashboards e relatórios** com volume por classe processual, ranking de relatores e autores citados.
  - Implementar um **recorte inicial de 6 meses**, servindo de MVP para validação técnica e de valor.
- 

## 5. Funcionalidades e Operação Técnica

A plataforma contempla um conjunto integrado de recursos que cobre todo o ciclo de ingestão, processamento e disponibilização de decisões do STF:

- **Fontes de Jurisprudência:** integração de uma **base histórica** via scraping retroativo e captura de **novas decisões** por monitoramento contínuo.
  - **Monitoramento:** periodicidade configurável (diário / semanal / quinzenal), com identificação, análise e catalogação automática, incluindo a coleta do **inteiro teor** (PDF/HTML) e anexos.
  - **Coleta e Preservação:** arquivamento permanente dos originais, aplicação de **OCR** quando necessário e garantia de **rastreabilidade**.
  - **Estruturação de Metadados:** organização em modelo relacional confiável com campos essenciais (processo, classe, relator, órgão, datas, partes).
  - **Citações Doutrinárias:** identificação e normalização de autores, obras e artigos, integrando entidades de **decisão**  $\rightleftharpoons$  **citação**  $\rightleftharpoons$  **obra**  $\rightleftharpoons$  **autor**.
  - **Busca e Dashboards:** indexação full-text com filtros avançados, séries temporais, rankings e indicadores estatísticos, além de relatórios exportáveis.
  - **Pipeline Técnico (ETL):** etapas de coleta, parsing, extração de metadados, validação, inserção, geração de logs de qualidade, indexação e disponibilização em dashboards. Fonte primária: Portal do STF (jurisprudência e inteiro teor).
  - **Scraping Retroativo:** implementação inicial cobrindo os últimos 6 meses (ex.: 01/03/2025 a 31/08/2025), com **cache local** para reprocessamento eficiente.
- 

## 6. Modelo de Dados

**Entidades:** Decisão, Processo, Classe, Relator, Órgão, Citação, Obra, Autor.

**Relações:** - Decisão → Processo (1:1) - Decisão → Classe/Relator/Órgão (N:1) - Decisão → Citação (1:N) - Citação → Obra (N:1) - Obra → Autor (N:1)

---

## 7. Critérios de Aceitação

- Ingestão e indexação de **100%** das decisões no período.
  - **Consulta funcional** por processo e relator.
  - Dashboard exibe volume por classe, ranking de relatores e autores citados.
  - Relatórios básicos disponíveis.
  - **Trilha de auditoria** com originais e metadados de coleta.
- 

## 8. Roadmap

- **Fase 0:** Setup inicial.
  - **Fase 1:** MVP com ETL, indexação e dashboards.
  - **Fase 2:** NLP e normalização avançada.
  - **Fase 3:** Expansão para outros tribunais.
-