

## **Projeto Aplicado II**

# **TRANSPARÊNCIA SIMPLIFICADA**

**Cainã Fernandes Guimarães da Silva**  
**Gabriel de Oliveira Meloni**  
**Humberto Gutttau Bravo**  
**Jimmy Paiva Gomes**

- **Sobre o CAGE**

- Vinculada à Secretaria da Fazenda (SEFAZ), a CAGE é o órgão responsável por coordenar e executar as atividades de controle interno do Poder Executivo Estadual, atuando como um pilar para a boa governança e a gestão fiscal responsável.

- **Principais atribuições:**

- **Contabilidade:** Normatiza procedimentos contábeis, consolida contas públicas e elabora o Balanço Geral do Estado.
- **Auditoria e Controle:** Verifica legalidade, legitimidade e eficiência dos atos de gestão.
- **Transparência:** Garante acesso às informações públicas por meio do Portal da Transparência RS.
- **Integridade e Correição:** Prevê e combate irregularidades, promovendo ética e integridade no serviço público.



# Caracterização do Problema

---

## **CAGE e o desafio da acessibilidade:**

Apesar do Portal Transparência RS disponibilizar um grande volume de informações, documentos essenciais como o Balanço Geral do Estado são escritos em linguagem técnica (contábil e administrativa), sendo de difícil compreensão para a população em geral.

Isso cria uma barreira: os dados são públicos, mas não são acessíveis. Como consequência, o alcance da transparência é limitado e o efetivo controle social pela sociedade civil fica comprometido

**Resumindo:** De que adianta a transparência se o cidadão comum não consegue entender os dados para fiscalizar a gestão pública?



# Objetivo

---



O objetivo deste projeto é desenvolver um ChatBOT com IA para "traduzir" dados públicos complexos, como o Balanço Geral do Estado, para o cidadão. Para isso, é essencial extrair e processar o conteúdo dos documentos oficiais (textos, tabelas e gráficos) e desenvolver um modelo capaz de responder a consultas em linguagem simples.

## A proposta inclui as seguintes etapas:

- **Extração e Vetorização de Textos:** Identificar e processar os conteúdos textuais dos documentos, simplificando termos técnicos para criar uma base de conhecimento compreensível.
  - **Extração e Organização de Dados Visuais:** Sistematizar informações contidas em gráficos e tabelas, criando um banco de dados estruturado que permita à IA interpretar elementos visuais.
  - **Desenvolvimento do ChatBOT com IA Generativa:** Implementar o modelo de IA capaz de processar as consultas, correlacionar dados textuais e visuais, e fornecer respostas precisas e contextualizadas em linguagem acessível.
- 
-

# Análise Exploratória - Metodologia

---

## 1. Identificação da Base de Dados (O Quê?)

- **Fonte Principal: O Balanço Geral do Estado (BGE) de 2024.**
  - Documento técnico de 174 páginas, publicado em 2025.
  - Detalha a execução financeira e os desafios fiscais do RS.

## 2. Definição da Linguagem de Programação (O Como?)

- **Linguagem Escolhida: Python**
- **Por que Python?**
  - **Aderência à IA:** É a linguagem de referência para Inteligência Artificial, Machine Learning e Processamento de Linguagem Natural (NLP).
  - **Ecossistema de Bibliotecas:** Possui ferramentas robustas para extração de dados (de PDFs, tabelas) e uso de modelos de linguagem.
  - **Versatilidade:** Permite o desenvolvimento completo da solução (análise, backend e APIs).



# Análise Exploratória - Conceitos

---



## 1. Processamento de Dados Não Estruturados :

- **Conceito:** Documentos (PDFs, textos, imagens) não possuem formato rígido.
- **Aplicação:** Uso de **OCR** (via biblioteca *Docling*) para "ler" e extrair todo o conteúdo bruto (textos e tabelas) do Balanço Geral do Estado.

## 2. Engenharia de Dados

- **Conceito:** Construção de fluxos de dados (ETL).
- **Aplicação:** Uso de **LangChain** para criar um pipeline que processa os dados extraídos, limpa-os e os transforma em um formato utilizável.

## 3. Tópicos de Banco de Dados:

- **Conceito:** Armazenamento e recuperação eficiente de informação.
- **Aplicação:** Os textos processados são convertidos em Embeddings (vetores numéricos) e armazenados em uma Vector Store (um banco de dados NoSQL otimizado para busca de similaridade).

## 4. Inteligência Artificial

- **Conceito:** NLP (Processamento de Linguagem Natural) e IA Generativa.
- **Aplicação:**
  1. A pergunta do usuário é convertida em um vetor.
  2. O sistema faz uma busca semântica (NLP) na *Vector Store* para encontrar os trechos mais relevantes do Balanço.
  3. A IA Generativa (LLM) lê esses trechos e "traduz" a resposta técnica para uma linguagem simples e objetiva

# Análise Exploratória - Acurácia

---

## Como Medimos o Sucesso? (Nossa "Acurácia")

Nosso objetivo é medir o quanto um cidadão comum entende a resposta. Para isso, criamos uma métrica chamada "Compreensividade Artificial". que as "personas" deram.

### O cálculo é feito em 3 passos:

1. **Avaliadores Virtuais:**
  - Usamos IAs que simulam "personas" (cidadãos leigos, como uma enfermeira ou estudante) para ler as respostas do chatbot.
2. **Sistema de Notas (de 1 a 5):**
  - Cada "persona" dá uma nota para a resposta:
  - **Nota 1:** Não entendeu nada (muito técnico).
  - **Nota 5:** Entendeu tudo (texto claro e simples).
3. **Resultado Final (A Média):**
  - A nossa "nota de Compreensividade" é simplesmente a **média** de todas as notas que as "personas" deram.

### Justificativa

- Alinhar-se ao objetivo do projeto de tornar a informação clara e acessível;
- Ser viável e de baixo custo, permitindo avaliação automatizada
- Fornecer indicadores preliminares de qualidade antes de testes com usuários;
- Possibilitar evolução futura para avaliações com cidadãos, ampliando a representatividade.

# Análise Exploratória - Acurácia

---



Fórmulas:

$$\text{Compreensividade} = \frac{\text{Quantidade de respostas compreendidas pelos usuários}}{\text{Total de respostas avaliadas}}$$

$$\text{Compreensividade Artificial} = \frac{\sum_{i=1}^n \text{nota}_i}{n}$$





# Resultados – Metodologia do Teste

---



Para obter os resultados de acurácia e desempenho do modelo, foi realizado um teste simulado que avaliou a clareza e a acessibilidade das respostas do ChatBOT.

## 1. Estrutura do Teste

- **Dataset:** 50 perguntas elaboradas a partir de dados do **Balanco Geral do Estado**.
- **Avaliação:** Respostas do ChatBOT avaliadas por um **LLM** que simulou **5 personas** distintas.
- **Métrica:** Cada persona atribuiu nota de **1 (Nenhuma Compreensão)** a **5 (Compreensão Plena)**.

## 2. Personas Simuladas (10 perguntas por persona)

- 🧑 Cidadão Leigo (Aposentado)
- 📰 Jornalista Local
- 🎓 Estudante Universitário
- 📊 Contador Público
- 🏛 Vereador Municipal

# Resultados - Análise

## 1. Resultados Agregados (N=50):

- **Média Geral: 3,86** (de 5,0)
  - *Interpretação: Desempenho entre "Compreensão Parcial" (3) e "Boa Compreensão" (4).*
- **Distribuição das Notas:**
  - **Nota 5 (Plena):** 1 resposta (2%)
  - **Nota 4 (Boa):** 41 respostas (82%)
  - **Nota 3 (Parcial):** 8 respostas (16%)
  - **Nota 2 ou 1:** 0 respostas (0%)

## 2. Análise dos Dados:

- **Ponto Forte:** 84% das respostas (42 de 50) foram consideradas de "Boa" a "Plena" compreensão.
- **Ponto de Destaque:** 0% de notas 1 ou 2. O modelo evitou com sucesso respostas incompreensíveis ou com excesso de jargão técnico.
- **Oportunidade:** 16% (Nota 3) indicam que a ideia central foi entendida, mas com lacunas, apontando para refinamentos futuros.

