

# Dataset para avaliação de sumarização de longos documentos em português

Leandro Carísio Fernandes  
Guilherme Zeferino Rodrigues Dobins

# Descrição do Projeto

Criação de dataset contendo:

- (TEXTO): transcrição de audiências públicas realizadas na Câmara dos Deputados
- (SUMÁRIO): notícias sobre essas audiências
- (SUMÁRIO ESTRUTURADO): metadados extraídos das notícias indicando opiniões:quem disse/apoia o quê

# Descrição do Projeto

Extração das principais opiniões a partir das transcrições

- Utilizamos o ChatGPT para extrair as principais opiniões a partir das transcrições.
- Inicialmente, a ideia era também utilizar o Gemini para a extração, mas ele não produziu respostas aceitáveis, logo descartamos a ideia
- Após as extrações das opiniões, comparamos com o dataset ouro para avaliar os resultados
- Por fim, avaliamos possíveis alucinações na extração de opiniões do GPT.

# Avaliação da extração de opiniões

- Após a extração das principais opiniões pelo ChatGPT, criamos um pipeline de avaliação automática para verificar a conformidade das extrações em relação ao dataset ouro
- O pipeline consiste nos seguintes passos:
  1. Para cada caso (audiência), mapeamos os indivíduos presentes no dataset ouro e os presentes na extração.
  2. Para cada indivíduo mapeado, enviamos as opiniões presentes no dataset ouro e as opiniões extraídas para o ChatGPT para um LLM, para que ele faça um mapeamento das opiniões.
  3. A partir do mapeamento, calculamos a quantidade de opiniões presentes no dataset ouro que foram escritas pelo extrator, e vice-versa, obtendo assim a precisão e o recall
- Recall médio: 0.45
- Precisão média: 0.25

# Detecção de alucinações

Além de verificarmos a precisão e o recall comparando as opiniões extraídas com as presentes no dataset ouro, foi necessário verificar se as opiniões extraídas eram realmente válidas, ou apenas alucinações do ChatGPT, visto que elas foram obtidas de transcrições muito extensas.

O pipeline de detecção de alucinações foi o seguinte:

1. Segmentação de todas as transcrições, isolando as falas de cada indivíduo presente na audiência.
2. Criação de índices de busca para cada indivíduo de cada audiência, para que possamos buscar as falas mais relevantes para alguma opinião.
3. Para avaliar se uma opinião é fruto de alucinação, buscamos os trechos mais relevantes no índice específico, enviamos tais trechos para um LLM junto da opinião, e o LLM detecta se é possível inferir a opinião a partir dos trechos.

# Segmentação das transcrições

```
"O SR. PRESIDENTE(Pinheirinho. Bloco/PP - MG)": [  
    "Bom dia a todos. Muito obrigado a cada um dos s  
    "Muito obrigado, Dra. Monize Marques, pelas impo  
    "Muito obrigado.\n\nAgora participará de forma v  
    "Muito obrigado, Sra. Patrícia Moreira.\n\nVamos  
    "A tecnologia nos ajuda muito, mas muitas vezes  
    "Eu também estou ficando velho. Um dado interess  
    "Muito obrigado, Dra. Monize.\n\nA Dra. Olinda c  
    "Muito obrigado, Dra. Olinda. Pudemos ouvir com  
    "Muito obrigado, Sr. Kenio.\n\nEu vou passar ago  
    "Muito obrigado, Sra. Patrícia.\n\nAgora, para a  
    "Muito obrigado, Sra. Monize.\n\nTem a palavra a  
    "Muito obrigado, Dra. Olinda.\n\nAgradeço, mais  
    ],  
    "A SRA. MONIZE DA SILVA FREITAS MARQUES": [  
        "Bom dia. Cumprimento a todos os presentes e tam  
        "Eu vou aproveitar a oportunidade, então, para t  
        "Dr. Kenio, Dra. Olinda e Dra. Patrícia, muito c  
    ],  
    "O SR. KENIO COSTA DE LIMA": [  
        "Bom dia a todas as pessoas que estão presentes  
        "Obrigado mais uma vez, Deputado.\n\nOuvi atenta  
    ],  
    "A SRA. PATRICIA MOREIRA": [  
        "Bom dia a todos, bom dia Sr. Kênio Costa, Exma.  
        "V.Exa. está muito longe disso.",  
        "Diante do que vocês falaram, eu concordo com tu  
    ],  
    "A SRA. OLINDA VICENTE MOREIRA": [  
        "Bom dia a todos e todas. Eu os cumprimento, rep  
        "Muito obrigada, Deputado.\n\nNo sentido da fala  
    ]  
]
```

# Identificação do índice correto

Nomes:

```
Nome: Bia Kicis. Cargo: Presidente da Comissão, D  
Nome: Eder Gatti. Cargo: Diretor do Departamento  
Nome: Francisco Eduardo Cardoso Alves. Cargo: Méd  
Nome: Caio Salvino. Cargo: Farmacêutico bioquímico  
Nome: Roberto Zeballos. Cargo: Médico, mestre e c
```

Índices::

```
O SR FRANCISCO EDUARDO CARDOSO ALVES  
O SR CORONEL CHRISÓSTOMO PL RO_  
O SR TADEU VENERI Bloco PT PR_  
O SR CAIO SALVINO_  
O SR GENERAL GIRÃO PL RN_  
O SR ROBERTO ZEBALLOS_  
O SR JORGE SÖLLA Bloco PT BA_  
O SR SARGENTO GONÇALVES PL RN_  
O SR PRESIDENTE André Fernandes PL CE_  
O SR OSMAR TERRA Bloco MDB RS_  
A SRÁ PRESIDENTE Bia Kicis PL DF_  
O SR DR FREDERICO Bloco PATRIOTA MG_  
O SR EDER GATTI_  
O SR DELEGADO PAULO BILYNSKYJ PL SP_  
A SRÁ FRANCIANE BAYER Bloco REPUBLICANOS RS_  
A SRÁ
```

# Identificação do índice correto

## Mapeamento dos índices

```
{  
  'Bia Kicis': 'audiencia_15_A_SRA_PRESIDENTE_Bia_Kicis_PL_DF_',  
  'Eder Gatti': 'audiencia_15_O_SR_EDER_GATTI',  
  'Francisco Eduardo Cardoso Alves': 'audiencia_15_O_SR_FRANCISCO_EDUARDO_CARDOSO_ALVES',  
  'Caio Salvino': 'audiencia_15_O_SR_CAI0_SALVINO',  
  'Roberto Zeballos': 'audiencia_15_O_SR_ROBERTO_ZEBALLOS'  
}
```



# Busca dos trechos mais relevantes para a opinião

- Uma opinião é passada para o buscador, no índice correto, e os 3 chunks mais relevantes são retornados.

Opinião: Destaca a importância da educação, saúde e oportunidades no mercado de trabalho para pessoas com TEA

Trecho mais relevante: Acredito que é responsabilidade do Estado assegurar que todos tenham acesso à educação de qualidade, aos serviços de saúde adequados e à oportunidade no mercado de trabalho, independentemente de suas diferenças. No entanto, sabemos que ainda há desafios a superar. A inclusão das pessoas com autismo requer não apenas leis políticas, mas também uma mudança profunda em nossa cultura e mentalidade. Precisamos construir uma sociedade mais empática, acolhedora e inclusiva, na qual cada indivíduo seja valorizado pelo que é...

# Detecção da validade da opinião

Finalmente, a opinião é passada para o LLM, seguida pelos trechos retornados da busca, e o LLM retorna algo no seguinte formato:

```
{  
  'explicacao': 'A opinião apresentada na matéria destaca a importância da educação, saúde e oportunidades no  
mercado de trabalho para pessoas com TEA. Os trechos da audiência fornecem evidências claras de que o indivíduo  
valoriza esses aspectos. Em particular, os trechos mencionam a responsabilidade do Estado em assegurar acesso à  
educação de qualidade, serviços de saúde adequados e oportunidades no mercado de trabalho para todos, incluindo  
pessoas com autismo. Além disso, há menções específicas sobre a necessidade de centros de atenção ao autismo e a  
luta por mais inclusão e atendimento. Portanto, é possível inferir que o indivíduo tem a opinião destacada na  
matéria.',  
  'resposta': 'True'  
}
```

# Detecção da validade da opinião

A resposta final pode ser “True”, caso a opinião seja válida, ou “False”, caso o LLM detecte que não é possível inferir a opinião a partir dos trechos da busca.

Ao fim, armazenamos a avaliação de cada opinião de cada indivíduo, e contamos a proporção de alucinação em cada caso.

Resultados nos 55 primeiros casos:

- Total opiniões válidas: 1008
- Total opiniões alucinadas: 87
- Proporção de alucinações: 0.079

# Dificuldades encontradas - Identificação do índice

- Os nomes extraídos pelo ChatGPT podem ser bem diferentes dos nomes que aparecem na transcrição.

e.g. “A SRA. LOURDES CARDOZO LAUREANO” aparece apenas como “Lourdinha”

- Uma mesma pessoa pode aparecer com mais de um nome na transcrição.

e.g.

“O SR PRESIDENTE Osmar Terra Bloco MDB/RS” = “O SR OSMAR TERRA Bloco MDB/RS”, ou

“O SR SEBASTIÃO COELHO” = “O SR SEBASTIÃO COELHO”

- Em certos casos, o GPT extraiu opiniões de pessoas que não estavam presentes na audiência, apenas foram citadas.

# Dificuldades encontradas - Limitações dos LLMs

Em muitos casos, um LLM falhava no momento de identificar se era possível inferir uma opinião a partir dos trechos.

Opinião:

Resposta GPT-3.5:

Resposta GPT-4: {

'explicacao': 'Os trechos fornecidos não mencionam diretamente a importância dos Estados Unidos para a liberdade de expressão. Eles discutem questões relacionadas à mentira, ameaças, e a remoção de conteúdo em várias jurisdições, incluindo os Estados Unidos, mas não destacam especificamente a importância dos Estados Unidos nesse contexto. Portanto, não é possível inferir a opinião apresentada na matéria a partir dos trechos fornecidos.'

'resposta': 'False'

}

# Dificuldades encontradas - Limitações dos LLMs

Em muitos casos, um LLM falhava no momento de identificar se era possível inferir uma opinião a partir dos trechos.

Opinião: Destacou a importância dos Estados Unidos para a liberdade de expressão.

GPT 3.5:

```
{  
  'explicacao': 'Nos trechos da fala, o indivíduo menciona a importância da liberdade de expressão ao falar sobre a repercussão do trabalho do jornalista Michael Shellenberger e a resistência em defesa da democracia. Além disso, ele critica aqueles que mentem para desmoralizar aqueles que defendem a liberdade. Portanto, é possível inferir que o indivíduo tem a opinião destacada na matéria.',  
  'resposta': 'True'  
}
```

GPT 4:

```
{  
  'explicacao': 'Os trechos fornecidos não mencionam diretamente a importância dos Estados Unidos para a liberdade de expressão. Eles discutem questões relacionadas à mentira, ameaças, e a remoção de conteúdo em várias jurisdições, incluindo os Estados Unidos, mas não destacam especificamente a importância dos Estados Unidos nesse contexto. Portanto, não é possível inferir a opinião apresentada na matéria a partir dos trechos fornecidos.',  
  'resposta': 'False'  
}
```

# Possibilidades futuras

- Além da tarefa de extrair opiniões da transcrição, seria possível avaliar a tarefa de gerar uma matéria a partir das opiniões do dataset ouro, ou escrever uma matéria diretamente da transcrição.
- Seria possível realizar as extrações com novos LLMs, para avaliar suas capacidades de sumarização