

# Atividade 01 - Definição do tema

## Equipe

**Nome:** *DataMinds*

**Repositório GitLab da equipe:** <https://gitlab.com/EricFukuyama/IntroCD2>

**Membros:**

- Anderson Nogueira Silva, 2126516, Andxyz8, andxyz8, [andersonnogueira@alunos.utfpr.edu.br](mailto:andersonnogueira@alunos.utfpr.edu.br), Eng. Comp., UTFPR
- Eric Yutaka Fukuyama, 2126567, EricFukuyama, ericfukuyama, [ericfukuyama@alunos.utfpr.edu.br](mailto:ericfukuyama@alunos.utfpr.edu.br), Eng. Comp., UTFPR
- Erick Jose Teles de Andrade, 2126575, @erickandrade1, Erick Andrade#3830, [erickandrade@alunos.utfpr.edu.br](mailto:erickandrade@alunos.utfpr.edu.br), Eng. Comp., UTFPR

## Tema

Análise dos discursos de diferentes nações no Debate Geral da Assembleia Geral da ONU.

## Objetivos

- Identificar quais tópicos globais foram mais discutidos por cada bloco (majoritariamente G7 e BRICS) em diferentes períodos.
- Analisar como mudanças geopolíticas, crises econômicas ou políticas impactam a retórica de cada grupo.
- Mapear os principais temas econômicos ao longo do tempo, destacando as semelhanças e contrastes para diferentes blocos econômicos.
- Analisar se os sentimentos de discursos de países aliados convergem ou divergem ao longo do tempo.

## Fontes de dados

**Corpus:** UN General Debate Corpus (UNGDC), coleção completa com mais de 10.000 discursos de 202 países, abrange os anos de 1946 a 2023. Última atualização ocorreu em agosto de 2024.

## Dados disponíveis

- **Discursos completos de países da ONU:** declarações e temas abordados por cada país durante o debate anual, os quais possuem tópicos de relevância política, econômica e social.
- **Metadados:** informações sobre o país, ano do discurso, número da sessão, orador, cargo ocupado pelo orador.

**Volume de dados:** possui 10.760 discursos no total, sendo que 857 discursos são de países do G7 e BRICS. Brasil, Estados Unidos, Canadá e Grã Bretanha possuem discursos em todas as seções e China, Índia e Rússia não possuem discurso em apenas uma das 78 sessões. Os outros países desses grupos possuem menos discursos. Os discursos, de modo geral, possuem em média de 2931 palavras, sendo que o discurso com mais palavras possui 22000 palavras e o menor 423.

**Uso dos dados:** os dados dos discursos serão utilizados visando extrair tópicos e padrões de retórica ao longo das décadas para diferentes países e grupos econômicos. Ainda, foi feita uma rápida análise da quantidade de discursos por países do G7 e do BRICS como demonstrado nas imagens abaixo.

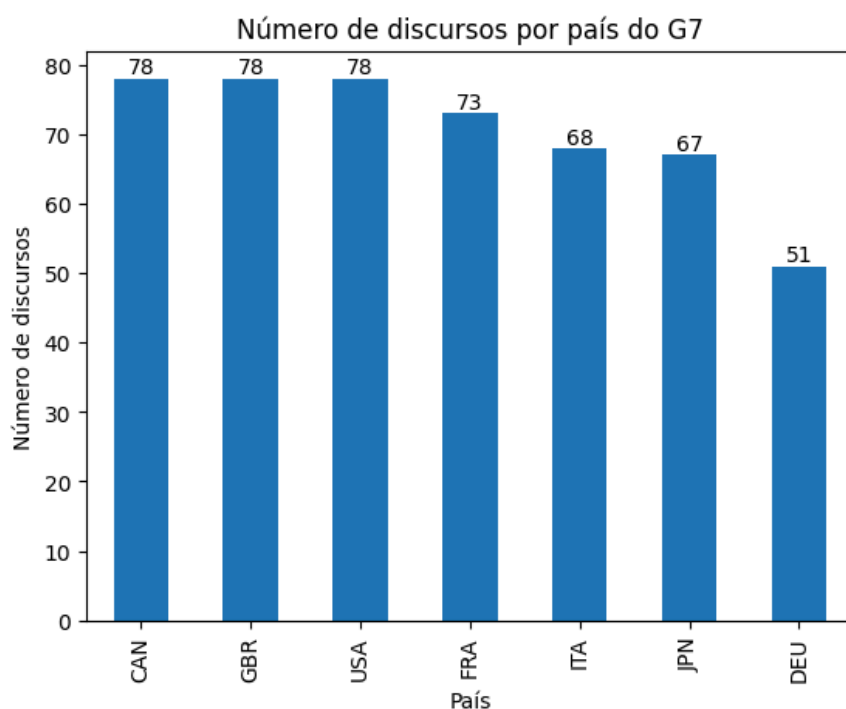


Figura 1 - Número de discursos por país do G7 presente no UNGDC.

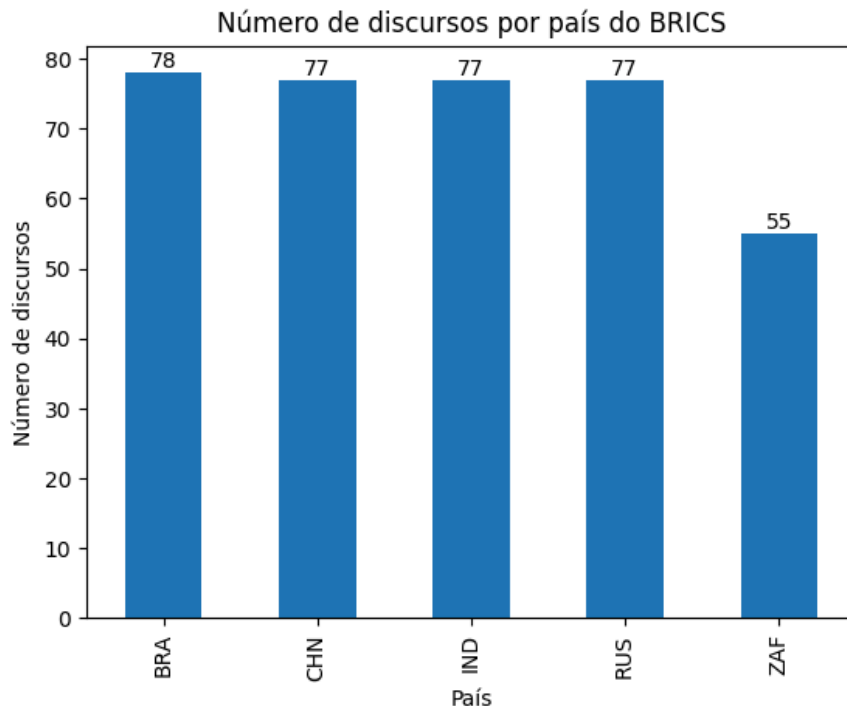


Figura 2 - Número de discursos por país do BRICS presente no UNGDC.

## Desafios

- **Interpretação geopolítica do contexto dos discursos:** a equipe considera que a análise desses temas exige um entendimento bem abrangente dos eventos históricos e o contexto geopolítico por trás que certamente influenciam os discursos. Portanto, para uma apresentação adequada das análises desenvolvidas será necessário o entendimento desses acontecimentos.
- **Período de análise longo:** o corpus UNGDC possui dados que abrangem os anos de 1946 a 2023, isto significa que pode ser necessário o conhecimento de contextos históricos de muitas décadas diferentes, o que pode fazer com que o entendimento dos dados analisados seja mais demorado. Uma alternativa seria dividir essa análise em períodos mais curtos (eg. períodos pré e pós Guerra Fria (1947 - 1991) e/ou em décadas anos 1990, 2000, 2010 etc).

## Referências

1. **HARVARD DATAVERSE.** UN General Debate Corpus (UNGDC), 1946-2023. Atualizado em 08/2024. Disponível em: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/0TJX8Y>. Acesso em: 24 out. 2024.