

# AI-Assisted Literature Review: Summarization using Transformer

**Presenter: Shabnam Barbhuiya**

# Seeing the Big Picture—Faster

## Traditional Method



Traditional literature reviews are **time-consuming, manual, and prone to bias.**

## Current Landscape



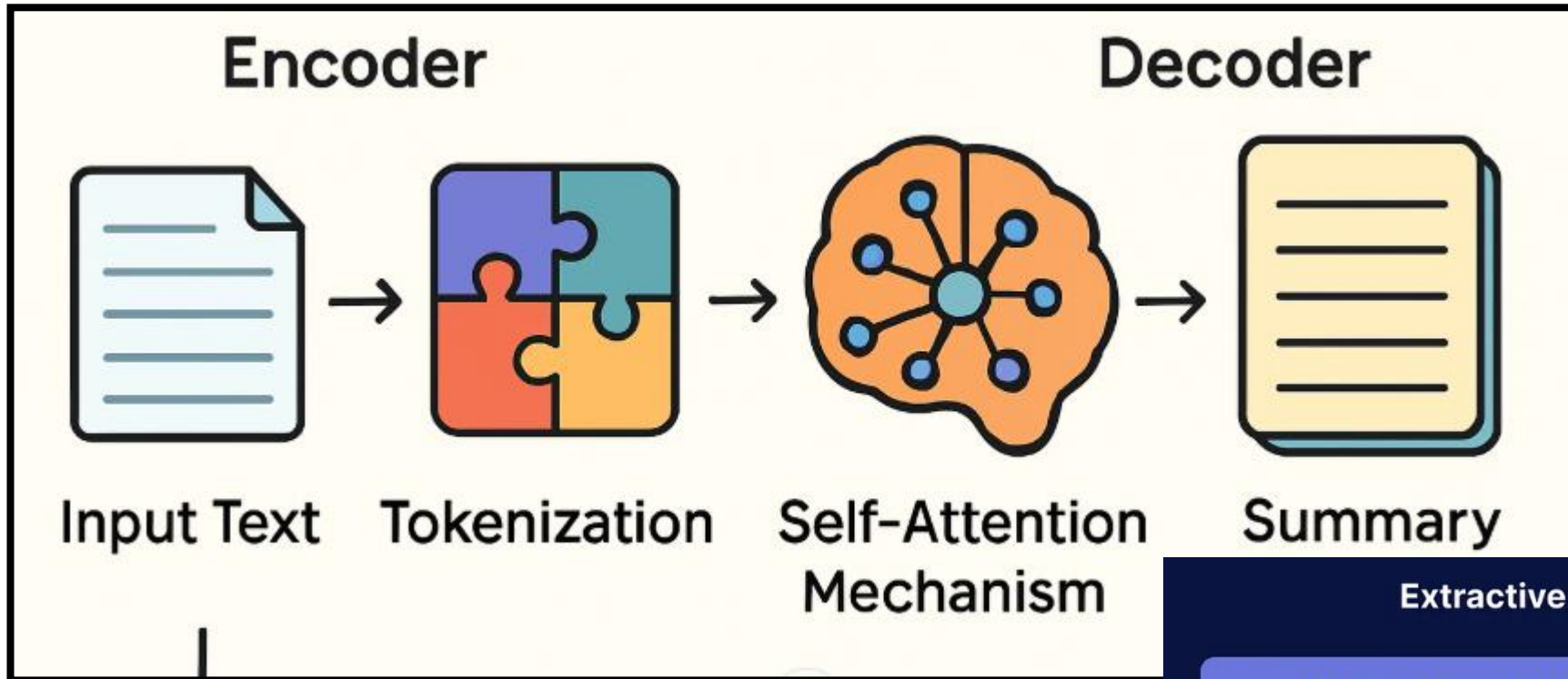
With the **exponential growth of publications**, it's harder to stay updated.



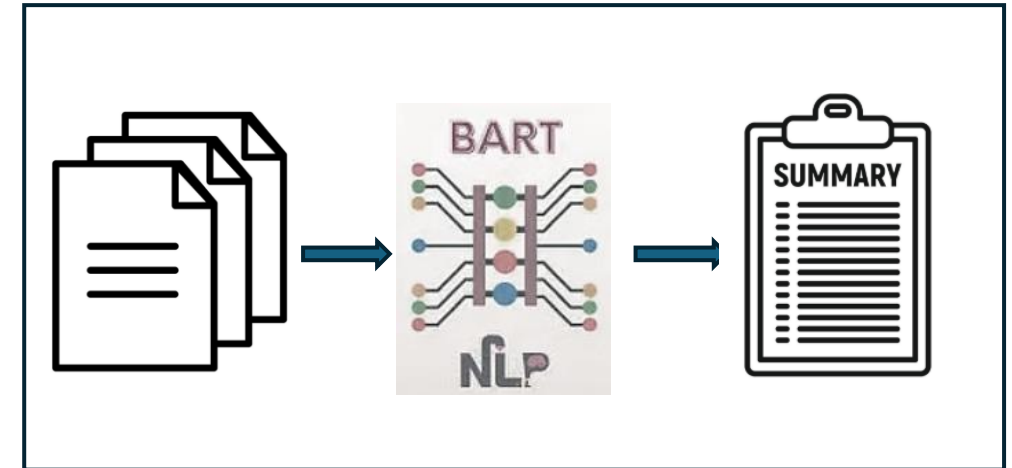
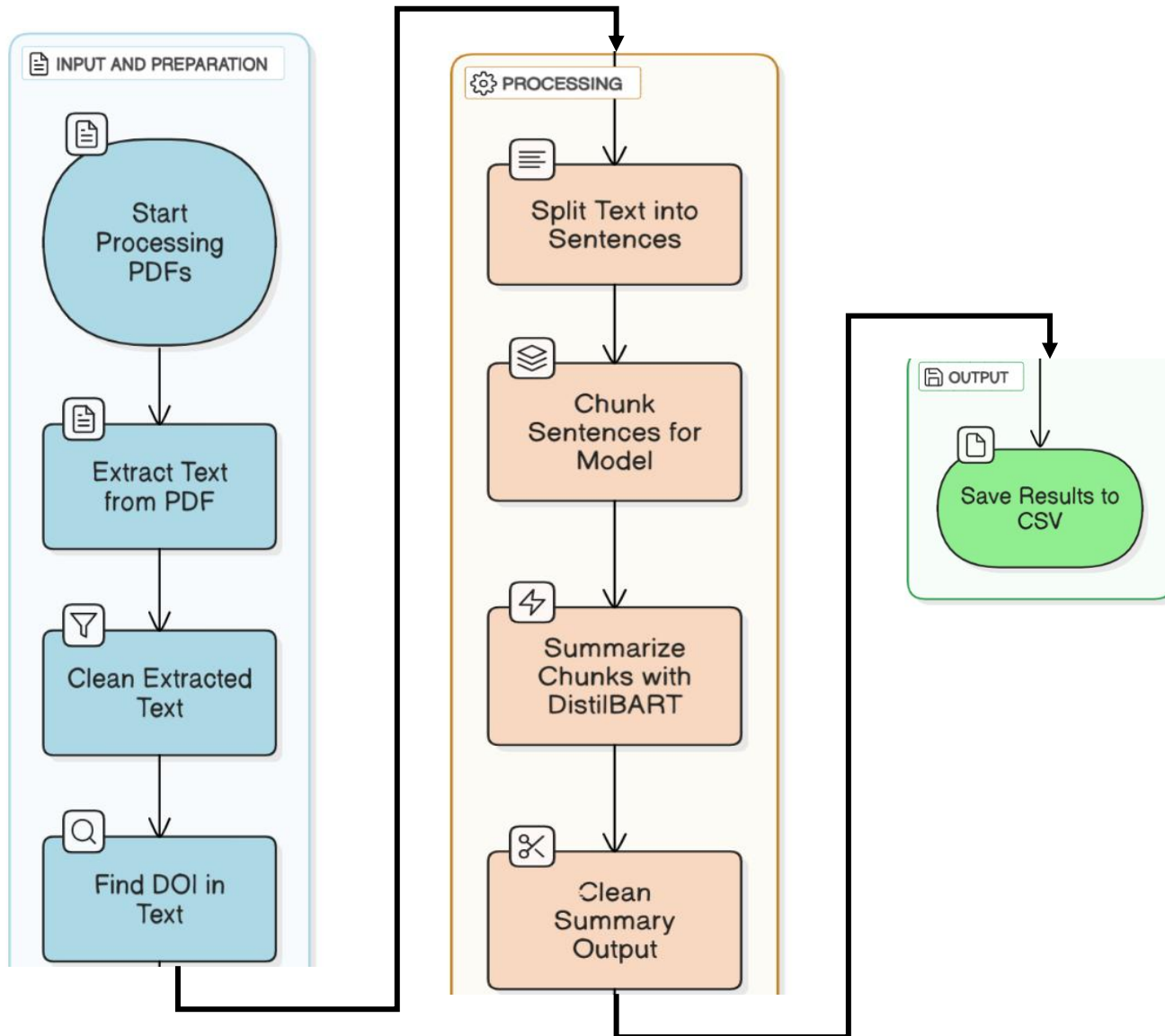
**AI and NLP technologies** now enable faster, scalable, and context-aware analysis



# Transformers at Work: Breaking Down Summarization



# Summarization Workflow



## Library Used:

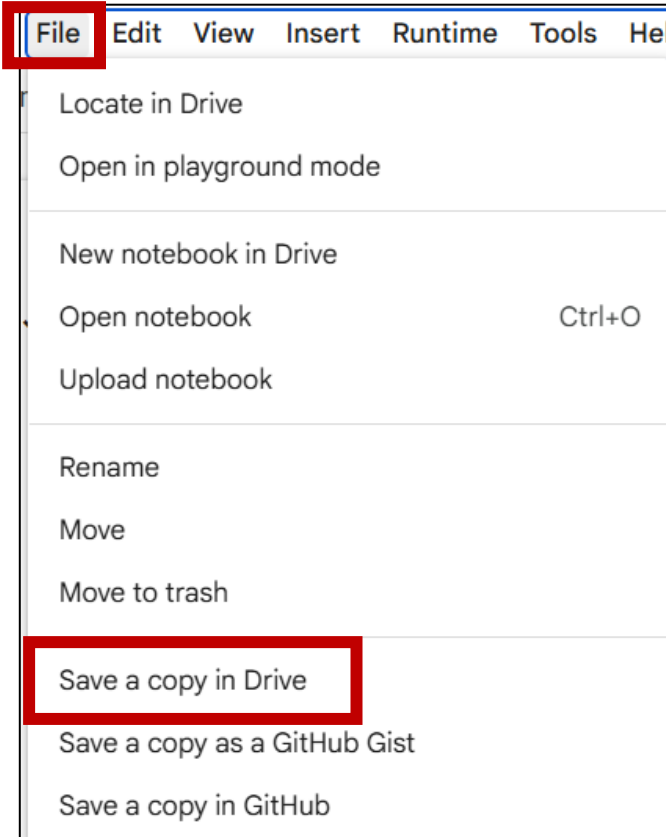
- **PyMuPDF** – Extracts clean text from PDFs
- **NLTK** – For sentence tokenization
- **DistilBART (via HuggingFace)** – Transformer model for summarization
- **Pandas** – Store DOI + Summary as CSV

**LET'S MOVE ON  
TO THE DEMO!**

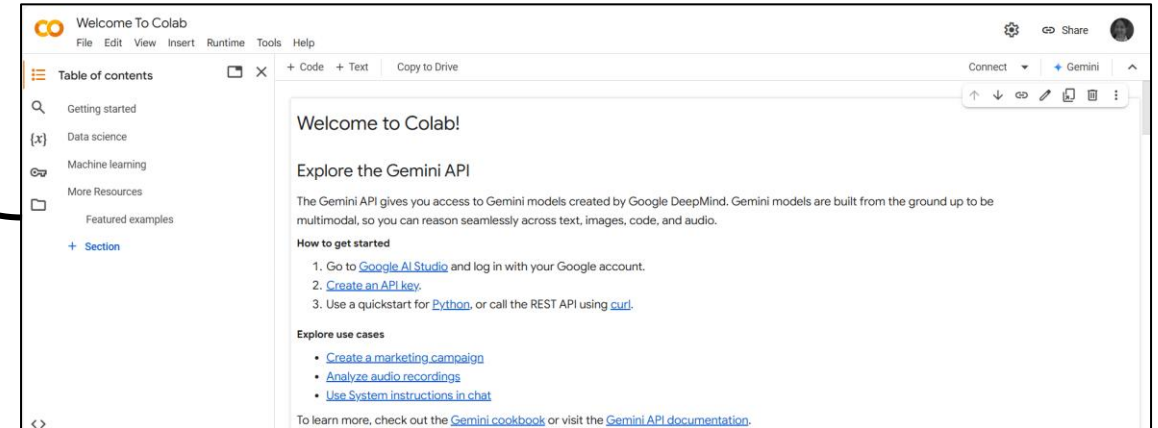


# INSTRUCTION

## Setting Up Your Colab Notebook



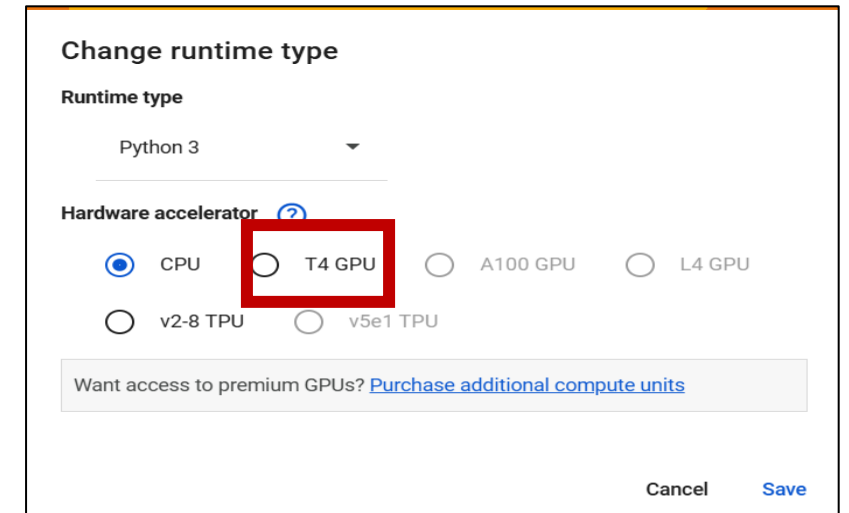
Home Page



**1. Copy the Colab Notebook to Your Own Google Drive** Go to the menu and click:

**File → Save a copy in Drive**

**2. Change the Runtime to GPU**  
In the Colab menu, select:  
**Runtime → Change runtime type**



# Step-by-Step Implementation in Colab

## Step 1 – Install Required Libraries

```
[1] !pip install transformers[sentencepiece] pymupdf nltk  
!pip install pygetpapers  
!pip install amilib==0.3.9
```

## Step 2 – Download Research Papers

Use 'pygetpapers' to search by topic/keyword and download PDFs.

```
!pygetpapers --query "biodiversity" --pdf --limit 5 --output downloaded_file --save_query
```

Search Query

Download fulltext PDFS

Save the passed query in a config file



# Step-by-Step Implementation in Colab

## Step 3 – Create and Visualize Retrieved Article using ‘amilib’

```
[ ] !amilib HTML --operation DATATABLES --indir downloaded_file
```

OUTPUT

pmcid	doi	title	authorString	journalInfo.journal.title	pubYear	abstractText
<a href="#">PMC12189379</a>	<a href="https://doi.org/10.3390/biology14060700">10.3390/biology14060700</a>	Marine Biodiversity Conservation Planning in the Indo-Pacific Convergence Zone Based on Ecological Spatial Analysis.	<ul style="list-style-type: none"><li>• Zhao L</li><li>• Li T</li><li>• Cong B</li><li>• Wang B</li><li>• Liu K</li><li>• Liu S.</li></ul>	Biology	2025	Marine biodiversity is of critical importance to global ecosystems. The Indo-Pacific Convergence Zone (IPCZ), a global marine biodiversity hotspot, faces escalating threats from human activities and c
<a href="#">PMC12212572</a>	<a href="https://doi.org/10.1371/journal.pgph.0004624">10.1371/journal.pgph.0004624</a>	Metrics for biodiversity and health policy integration.	<ul style="list-style-type: none"><li>• Nogués-Bravo D</li><li>• Whitmee S</li><li>• Willetts L.</li></ul>	PLOS global public health	2025	Despite over a decade of progressive commitments from parties to the Convention on Biological Diversity (CBD), integrated biodiversity and health indicators and monitoring mechanisms remain limited, h



# Step-by-Step Implementation in Colab

## Step 4 – Load PDFs for Summarization

```
import os

# USER INPUT: Either a single PDF file or a folder path
input_path = "/content/downloaded_file" # Replace with your own file or folder path
```

## Step 5 – Initialize Model & Tokenizer

```
# Download necessary NLTK resources
nltk.download('punkt')

# Load model and tokenizer from Huggingface
checkpoint = "sshleifer/distilbart-cnn-12-6" #facebook/bart-large-cnn #nsi319/legal-pegasus # google/pegasus-pubmed
tokenizer = AutoTokenizer.from_pretrained(checkpoint)
model = AutoModelForSeq2SeqLM.from_pretrained(checkpoint)
max_input_length = tokenizer.model_max_length
```

Default: `'sshleifer/distilbart-cnn-12-6'` (Changeable)

# Step-by-Step Implementation in Colab

## Step 6&7 – Summarize and Export

```
import nltk
nltk.download('punkt_tab')
summaries_dict = {}

for i, pdf_path in enumerate(pdf_paths, 1):
    print(f"Summarizing paper {i}/{len(pdf_paths)}: {os.path.basename(pdf_path)}")
    try:
        doi, summary = summarize_pdf(pdf_path) # Get both DOI and summary
        summaries_dict[pdf_path] = {"DOI": doi, "Summary": summary}
        print(f" DOI: {doi}\n Summary: {summary[:200]}...")
        #summaries_dict[pdf_path] = summary
```

Output

**PMCID:** PMC12212572

**DOI:** <https://doi.org/10.1371/journal.pgph.0004624>

► Click to view summary

**PMCID:** PMC12214654

**DOI:** <https://doi.org/10.1038/s41598-025-01959-3>

► Click to view summary

**PMCID:** PMC12241762

**DOI:** <https://doi.org/10.1111/gcb.70328>

► Click to view summary

**PMCID:** PMC12248781

**DOI:** <https://doi.org/10.3390/ani15131844>

► Click to view summary

**PMCID:** PMC12189379

**DOI:** <https://doi.org/10.3390/biology14060700>

► Click to view summary

# Output

ating Tree Species Traits for Biodive...

1 / 14

—

100%

+



animals



Article

## TreeGrid: A Spatial Planning Tool Integrating Tree Species Traits for Biodiversity Enhancement in Urban Landscapes

Shrey Rakholia <sup>1,†</sup> , Reuven Yosef <sup>2,\*</sup> , Neelesh Yadav <sup>1</sup> , Laura Karimloo <sup>3</sup> , Michaela Pleitner <sup>4</sup> and Ritvik Kothari <sup>1</sup>

<sup>1</sup> Bioinformatics Center, Forest Research Institute, Dehradun 248006, Uttarakhand, India; rakholias@gmail.com (S.R.); neeleesh\_yadav@icfre.org (N.Y.); ritvikkothari.cs2018@gmail.com (R.K.)

<sup>2</sup> Eilat Campus, Ben Gurion University of the Negev, P.O. Box 272, Eilat 881020, Israel

<sup>3</sup> Faculty of Forest and Environment, Eberswalde University for Sustainable Development, 16225 Eberswalde, Germany; laura.karimloo@hnee.de

<sup>4</sup> Centre for Microbiology and Environmental Systems Science, University of Vienna, 1030 Vienna, Austria; michaela.pleitner@gmail.com

\* Correspondence: ryosef60@gmail.com

† Share equal credit.

PMCID: PMC12248781

DOI: <https://doi.org/10.3390/ani15131844>

▼ Click to view summary

Shrey Rakholia, Reuven Yosef, Neelesh Yadav, Laura Karimloo, Michaela Pleitner and Ritvik Kothari write about TreeGrid: A Spatial Planning Tool Integrating The concept of wildlife-inclusive urban design involves integrating biodiversity as-pects into urban design at an early stage . Species-specific traits such as shade are particularly important in human-modified landscapes, where canopy cover enhances mammalian diversity and richness Rising LST has been shown to correlate with increased thermal stress and reduced precipitation in arid regions . Strategies such as sustainable land management and reforestation can help stabilize local climates and protect biodiversity Building upon recent advances in urban biodiversity Each tree species was assigned a faunal biodiversity score ranging from 0 to 5 . The scoring considered each species' capacity to provide food, shelter, and habitat resources . The connectivity index was computed by evaluating pairwise Haversine The training dataset comprised planted tree points with corresponding ecological attributes, including LST (°C), cooling index, shade index, height, and diameter . A binary label indicating the presence of birds was generated by checking whether "Bird Biodiversity and Ecology Metric Metric Range and Utility (Biodiversity Score (Trees) and ecosystem service scores (Ecosystem Services) Users have complete control over the tool, being able to toggle base layers (Imag The ROC curve for our DNN classifier (Figure 6) yields an AUC of , indicating good discrimination between points with and without bird presence . At most thresholds, the model achieves high specificity while maintaining a reasonable This study introduces an integrated spatial planning tool designed to tackle the issues of urban biodiversity loss, habitat fragmentation, and microclimate regulation through strategic tree placement . The tool can also be integrated into citizen science initiatives to expand biodiversity data collection and promote Study: Key lessons for biodiversity-sensitive cities and towns . Designing wildlife-inclusive cities that support human-animal co-existence . The influence of land surface temperature on Ghana's climate variability and implications for sustainable development . Private trees contribute uniquely to urban forest diversity, structure and service-based traits . Global patterns of diversity in the urban forest: Is there evidence to support the 10/20/30 rule? Urban For. Urban Green.

Before: ORIGINAL PAPER

After: SUMMARY

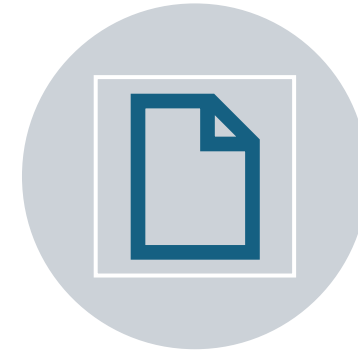
# !!!!Try It Yourself!!!!



- OPEN THE COLAB LINK PROVIDED AND  
MAKE YOUR OWN COPY.



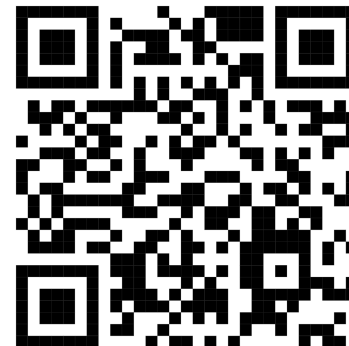
- PRACTICE UPLOADING, PROCESSING,  
AND SUMMARIZING A RESEARCH  
ARTICLE.



- EXPERIMENT WITH DIFFERENT PDFS  
AND SEE HOW THE SUMMARIES VARY.

**COLAB LINK:**

<https://colab.research.google.com/drive/1el5Zjogk7DXqqeuBzGMqFDBGTVyWg1Pm?usp=sharing>





# THANK YOU



**#semanticClimate**  
Transforming information into actionable knowledge